TABLE 29. POWER TABLE EXAMPLE

DB-05 TSA 85D - TARGET SPACING ANALYSIS (5 MILES, 950 FEET)

POWER TABLES FOR THE MAIN EFFECTS IN A 2X2 FACTORIAL REPEATED MEASURES ANOVA

ALPHA=.05

	BETA=.20				BETA=.10				BETA=.05				BETA=.01			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
N= 5	3.7	3.2	3.0	2.9	4.3	3.7	3.5	3.3	4.8	4.1	3.9	3.7	5.8	5.0	4.7	4.5
N= 6	3.2	2.7	2.5	2.4	3.7	3.1	2.9	2.8	4.1	3.5	3.3	3.1	4.9	4.2	3.9	3.8
N= 7	2.8	2.4	2.2	2.2	3.2	2.8	2.6	2.5	3.6	3.1	2.9	2.8	4.3	3.7	3.4	3.3
N= 8	2.5	2.2	2.0	2.0	2.9	2.5	2.3	2.3	3.2	2.8	2.6	2.5	3.9	3.3	3.1	3.0
N= 9	2.3	2.0	1.9	1.8	2.7	2.3	2.2	2.1	3.0	2.6	2.4	2.3	3.5	3.0	2.8	2.7
N=10	2.2	1.9	1.7	1.7	2.5	2.1	2.0	1.9	2.8	2.4	2.2	2.1	3.3	2.8	2.6	2.5
N=11	2.1	1.8	1.6	1.6	2.4	2.0	1.9	1.8	2.6	2.2	2.1	2.0	3.1	2.6	2.5	2.4
N=12	1.9	1.7	1.6	1.5	2.2	1.9	1.8	1.7	2.5	2.1	2.0	1.9	2.9	2.5	2.3	2.2
N=13	1.9	1.6	1.5	1.4	2.1	1.8	1.7	1.6	2.4	2.0	1.9	1.8	2.8	2.4	2.2	2.1
N=14	1.8	1.5	1.4	1.4	2.0	1.7	1.6	1.6	2.2	1.9	1.8	1.7	2.6	2.3	2.1	2.0
N=15	1.7	1.5	1.4	1.3	2.0	1.7	1.6	1.5	2.2	1.8	1.7	1.7	2.5	2.2	2.0	2.0
N=16	1.6	1.4	1.3	1.3	1.9	1.6	1.5	1.5	2.1	1.8	1.7	1.6	2.4	2.1	2.0	1.9
N=17	1.6	1.4	1.3	1.2	1.8	1.6	1.5	1.4	2.0	1.7	1.6	1.5	2.4	2.0	1.9	1.8
N=18	1.5	1.3	1.2	1.2	1.8	1.5	1.4	1.4	1.9	1.7	1.6	1.5	2.3	1.9	1.8	1.8
N=19	1.5	1.3	1.2	1.1	1.7	1.5	1.4	1.3	1.9	1.6	1.5	1.4	2.2	1.9	1.8	1.7
N=20	1.4	1.2	1.2	1.1	1.7	1.4	1.3	1.3	1.8	1.6	1.5	1.4	2.1	1.8	1.7	1.7

DB-05 TSA 85D - TARGET SPACING ANALYSIS (5 MILES, 950 FEET)

POWER TABLES FOR THE MAIN EFFECTS IN A 2X2 FACTORIAL REPEATED MEASURES ANOVA

ALPHA=.01

	BETA=.20				BETA=.10				BETA=.05				BETA=.01			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
N= 5	5.8	5.0	4.7	4.5	6.7	5.7	5.4	5.2	7.4	6.4	6.0	5.7	8.9	7.6	7.1	6.8
N= 6	4.7	4.0	3.7	3.6	5.3	4.6	4.3	4.1	5.9	5.0	4.7	4.5	7.0	5.9	5.6	5.4
N= 7	4.0	3.4	3.2	3.1	4.5	3.9	3.6	3.5	5.0	4.3	4.0	3.8	5.9	5.0	4.7	4.5
N= 8	3.5	3.0	2.8	2.7	4.0	3.4	3.2	3.1	4.4	3.8	3.5	3.4	5.1	4.4	4.1	4.0
N= 9	3.2	2.7	2.6	2.5	3.6	3.1	2.9	2.8	4.0	3.4	3.2	3.1	4.6	3.9	3.7	3.6
N=10	2.9	2.5	2.4	2.3	3.3	2.8	2.7	2.6	3.6	3.1	2.9	2.8	4.2	3.6	3.4	3.3
N=11	2.7	2.3	2.2	2.1	3.1	2.6	2.5	2.4	3.4	2.9	2.7	2.6	3.9	3.4	3.1	3.0
N=12	2.6	2.2	2.1	2.0	2.9	2.5	2.3	2.2	3.2	2.7	2.5	2.4	3.7	3.1	2.9	2.8
N=13	2.4	2.1	1.9	1.9	2.7	2.3	2.2	2.1	3.0	2.6	2.4	2.3	3.5	3.0	2.8	2.7
N=14	2.3	2.0	1.9	1.8	2.6	2.2	2.1	2.0	2.8	2.4	2.3	2.2	3.3	2.8	2.6	2.5
N=15	2.2	1.9	1.8	1.7	2.5	2.1	2.0	1.9	2.7	2.3	2.2	2.1	3.1	2.7	2.5	2.4
N=16	2.1	1.8	1.7	1.6	2.4	2.0	1.9	1.8	2.6	2.2	2.1	2.0	3.0	2.6	2.4	2.3
N=17	2.0	1.7	1.6	1.6	2.3	2.0	1.8	1.8	2.5	2.1	2.0	1.9	2.9	2.5	2.3	2.2
N=18	2.0	1.7	1.6	1.5	2.2	1.9	1.8	1.7	2.4	2.1	1.9	1.9	2.8	2.4	2.2	2.1
N=19	1.9	1.6	1.5	1.5	2.1	1.8	1.7	1.7	2.3	2.0	1.9	1.8	2.7	2.3	2.1	2.1
N=20	1.8	1.6	1.5	1.4	2.1	1.8	1.7	1.6	2.3	1.9	1.8	1.7	2.6	2.2	2.1	2.0

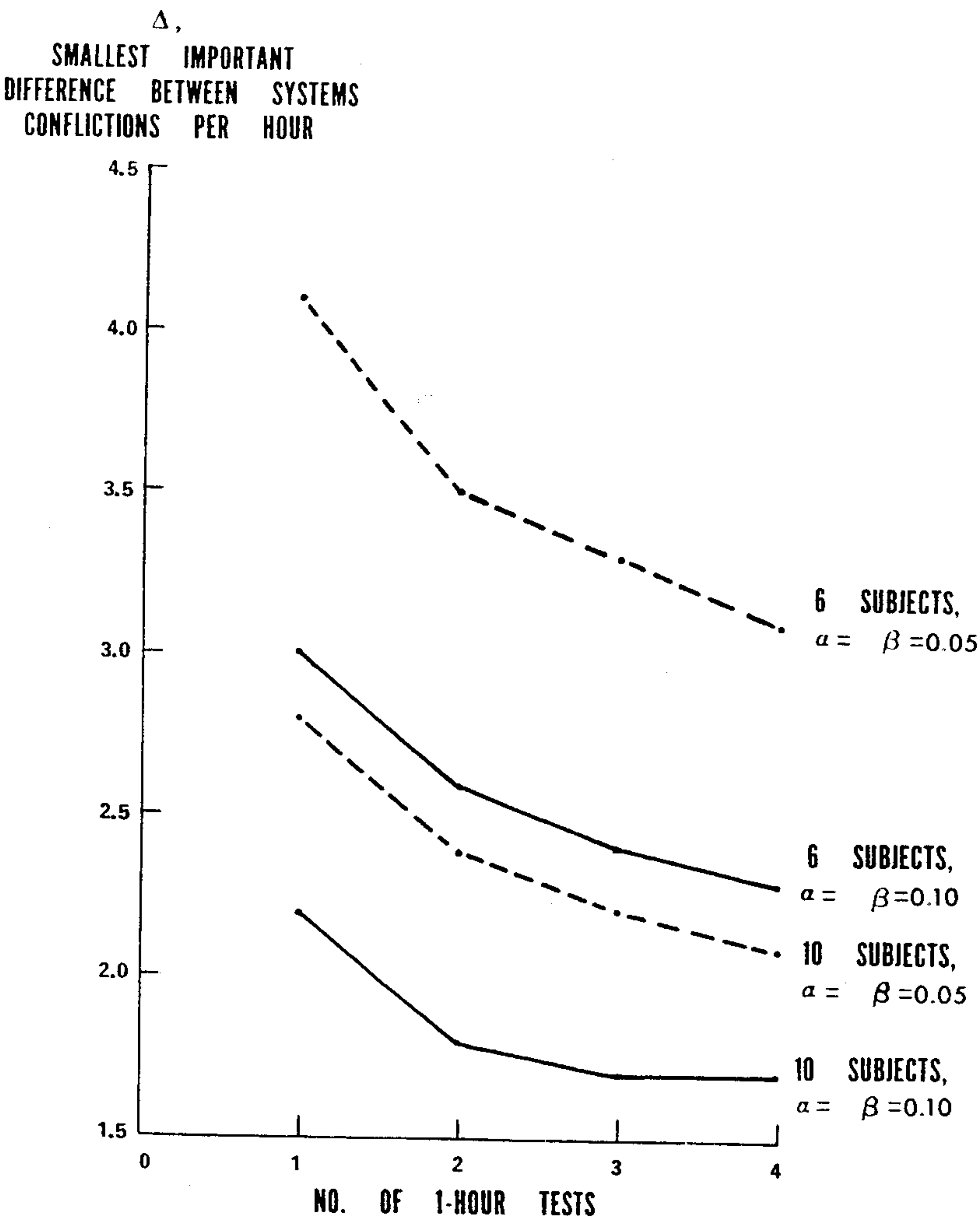

FIGURE 15. GRAPH OF POWER

TABLE 30. COMPARATIVE STATISTICAL POWER OF THE FOUR FACTOR SCORES

Paired t test (alpha=.10,beta=.01) one hour runs

NUMBER OF CASES	CONFLICT FACTOR	OCCUPANCY FACTOR	COMMUNICATION FACTOR	DELAY FACTOR
5	2.1	2.3	1.3	1.8
6	1.8	2.0	1.1	1.6
7	1.6	1.8	1.0	1.4
8	1.4	1.6	0.9	1.3
9	1.3	1.5	0.8	1.2
10	1.2	1.4	0.8	1.1
11	1.2	1.3	0.7	1.0
12	1.1	1.3	0.7	1.0
13	1.1	1.2	0.7	0.9
14	1.0	1.1	0.6	0.9
15	1.0	1.1	0.6	0.9
16	0.9	1.1	0.6	0.8
17	0.9	1.0	0.6	0.8
18	0.9	1.0	0.6	0.8
19	0.8	1.0	0.5	0.7
20	0.8	0.9	0.5	0.7
correlation *	.39	.44	.77	.07
standard dev. *	.7413	1.0296	.7742	0.3781

* based on running averages.

FOUR BASIC DESIGNS	Paired t Test 2 x 2 Repeated Measures ANOVA 2 x 3 Repeated Measures ANOVA Separate t Test	
TEN MEASURES	Confliction Factor Occupancy Factor Communication Factor Delay Factor	Conflictions (5 mi.) Aircraft Time Under Control Duration of Ground Air Com. Total Delay Time Number of Aircraft Handled Fuel Consumption Under Control
THREE DENSITY LEVELS	Low Medium High	
ALPHA ERROR LEVELS	.20 .10 .05 .01	
BETA ERROR LEVELS	.20 .10 .05 .01	
NUMBER OF SUBJECTS	6 to 20	
NUMBER OF REPLICATIONS	1 to 4	

DELTA (DETECTABLE INCREMENT) - in respective measures above

FIGURE 16. POWER TABLE STRUCTURE

These are cases in which important and expensive systems are tested, but because the power has not been adequately considered and thought about, the results which seem like clear improvements are found to be not significantly different from existing systems. This is likely if no allowance is made for the beta error and if the alpha level selected is too stringent for this purpose, leading to the erroneous finding of no significant difference.

AN EVALUATION OF THE INDEX OF ORDERLINESS

ANALYSIS. Frequently, new ideas for ATC system measures are suggested. It would be useful to have a method for evaluating such ideas. It is suggested here that a data base like the SEM data can be useful for this purpose. As an example of how that might be done, a brief examination is made of the measure "the index of orderliness" which had been omitted from the original list of measures. This measure was developed by Halvorsen at the FAA National Aviation Facilities Center (reference 10) and has been studied in various places, but has rarely been used in dyamic simulation studies of en route systems. It was examined as a way to evaluate air traffic control systems by Gent at the Royal Radar Establishment (RRE) (reference 11), and was applied in a U.S. Transportation Systems Center study (reference 12) cited by Horowitz in connection with his study of the ARTS III system (reference 13). The RRE thought it was a promising measure, and the Horowitz study group found it was highly related to time duration of the state of confliction.

As has been explained earlier, it was possible to re-score the basic data tapes containing the records of the simulation exercises. For scheduling reasons, it was decided to re-score only the SEM I data to obtain the index of orderliness for that experiment's runs. To be consonant with the other data from the simulation runs, it was necessary to develop some summary statistics to represent the run as a whole. Three such measures were generated. The basic form of the index of orderliness which was used and how the run scores were composed is discussed in detail in appendix E. The basic approach was to generate an index for each aircraft at each second of the problem, average these for the minute, and then average these over the hour. One of the three measures was this average, and another was the variance computed over the minutes for the hour, and the third was developed into what was called the "probablity expression of the index values." These will be referred to as "ORD 1," ORD 2," and "ORD 3."

Several criteria were used to evaluate these index of orderliness measures: the reliability of the three indexes, their correlations with other measures which might be expected to be similar, their correlations with the judges' ratings, and their multiple correlations with the judges ratings. As was mentioned above, Horowitz (reference 13) cited some work at TSC (reference 12) as indicating that there was a strong correlation with the confliction measures, notably the time two aircraft spent in a state of confliction, and the index of orderliness type of measure. This finding was confirmed. Table 31 presents the correlations for each of the six sector-density cells between the three versions of the index of orderliness and the four factor scores and the two major confliction measures, the number and duration of 5-mile (separation standard) conflictions. The correlations between the first two index of orderliness scores and two of the factor scores (confliction and occupancy) and the confliction measures are sometimes quite high, at least in one of the two sectors.

TABLE 31

CORRELATIONS BETWEEN INDEX OF ORDERLINESS MEASURES

AND FACTOR SCORES AND CONFLICTION MEASURES*

Sector 14, Density 1

	Ord 1	Ord 2	Ord 3
Confliction Factor	.44	.54	-.13
Occupancy Factor	.65	.29	.20
Communication Factor	.24	.09	.25
Delay Factor	.03	.28	.10
No. 5 Mile Conflicts	.34	.45	.12
Duration 5 Mile Conflicts	.36	.42	-.09

Sector 14, Density 2

	Ord 1	Ord 2	Ord 3
Confliction Factor	.79	.70	.08
Occupancy Factor	.77	.60	.19
Communication Factor	.29	.17	-.06
Delay Factor	.11	.13	.10
No. 5-Mile Conflicts	.78	.73	.08
Duration 5-Mile Conflicts	.77	.66	.13

Sector 14, Density 3

	Ord 1	Ord 2	Ord 3
Confliction Factor	.72	.78	.00
Occupancy Factor	.83	.77	.00
Communication Factor	.11	.01	.00
Delay Factor	-.42	-.28	.00
No. 5-Mile Conflicts	.55	.55	.00
Duration 5-Mile Conflicts	.78	.87	.00

CORRELATIONS BETWEEN INDEX OF ORDERLINESS MEASURES

AND FACTOR SCORES AND CONFLICTION MEASURES (CONTINUED)

Sector 16, Density 1

	Ord 1	Ord 2	Ord 3
Confliction Factor	.02	.19	-.01
Occupancy Factor	.23	.11	.37
Communication Factor	.12	.04	-.09
Delay Factor	-.16	-.08	-.03
No. 5-Mile Conflicts	-.15	.04	-.15
Duration 5-Mile Conflicts	.05	.29	.05

Sector 16, Density 2

	Ord 1	Ord 2	Ord 3
Confliction Factor	.38	.63	-.08
Occupancy Factor	.30	.19	.09
Communication Factor	-.37	-.29	+.24
Delay Factor	+.17	+.01	+.20
No. 5-Mile Conflicts	.36	.58	-.01
Duration 5-Mile Conflicts	.49	.46	.21

Sector 16, Density 3

	Ord 1	Ord 2	Ord 3
Confliction Factor	.27	.46	.00
Occupancy Factor	.52	.49	.00
Communication Factor	-.38	-.57	.00
Delay Factor	-.11	-.12	.00
No. 5-Mile Conflicts	.30	.43	.00
Duration 5-Mile Conflicts	.33	.56	.00

* Data based on two-run aggregates; N is generally 27-31.

Table 32 presents the correlations among the three index of orderliness scores for each of the six cells. The first two index of orderliness scores (ORD 1 and ORD 2 in the table) seem well correlated with each other, but ORD 3 seems only occasionally related to the others.

In table 33, the run-to-run reliabilities based on the correlations between two similar runs are shown. The reliability coefficients are shown for the index of orderliness variables in comparison to the four factor scores and the two conflict measures. The first two index of orderliness measures are not better than the other measures, and the third index of orderliness measure is somewhat worse. The general inadequacy of 1-hour runs as to reliability has been discussed earlier; in addition, it will be recalled that the SEM I runs were reduced by 10 minutes to adjust for computer data losses by maximizing the number of runs of the same length.

In table 34 are shown the relationships of these measures to the observer ratings. These are not remarkably stronger than others, and they differ somwhat in the two sectors.

Thus far, it is seen that the index of orderliness measures are highly correlated with each other, highly correlated with two of the four factor scores, and have nothing in particular to add in the way of reliabiliity. In one final analysis, let us examine them in the light of whether they can add anything to our already available prediction of the judges' ratings by the four factor scores. These multiple R's are shown in table 35, compared to the multiple R's found <u>without</u> these measures added in. The index of orderliness measures add very little.

The fact that these new measures add very little to the prediction of the judges' scores suggests that much of the variation these new measures carry is already accounted for by the four factor scores. If this is true, then perhaps the two factor scores which are most highly correlated with the indexes can, taken together, allow us to dispense with the index scores. Using this approach, the two factor scores for confliction and occupancy were averaged and the resulting average was correlated with the index meeasures. These correlations are shown in table 36.

As was just speculated, the two factor scores combined do account for a great deal of the two main index of orderliness measures' variance in several of the conditions studied, but again there is a marked difference in the correlations depending on the sector involved. This sector difference raises a question beyond the scope of the present exploration of the index of orderliness measures.

<u>IMPLICATIONS</u>. The index of orderliness measurement type seems to have some puzzling but interesting qualities. It is suggested that it is still worth further examination. Its examination here was not complete. The primary purpose of its examination here was to exemplify this method of using a data base to study measures other than those that had been included in the original study.

TABLE 32

CORRELATIONS AMONG THE THREE INDEX OF ORDERLINESS MEASURES*

		Density 1			Density 2			Density 3		
		ORD 1	ORD 2	ORD 3	ORD 1	ORD 2	ORD 3	ORD 1	ORD 2	ORD 3
Geom.1	ORD 1	1.00			1.00			1.00		
	ORD 2	.79	1.00		.89	1.00		.80	1.00	
	ORD 3	.45	.37	1.00	.21	.18	1.00	.00	.00	1.00
Geom.2	ORD 1	1.00			1.00			1.00		
	ORD 2	.89	1.00		.78	1.00		.69	1.00	
	ORD 3	.68	.54	1.00	.41	.31	1.00	.00	.00	1.00

*Data based on two-run aggregates; N is generally 27-31

TABLE 33

RUN-RUN* RELIABILITIES FOR INDEX OF ORDERLINESS

MEASURES, FACTOR SCORES AND CONFLICTION MEASURES

	Confl. Fac.Sc.	Occ. Fac.Sc.	Comm. Fac.Sc.	Delay Fac.Sc.	Confl. Count (5 mi.)	Confl. Dura.	ORD 1	ORD 2	ORD 3
G1 D1	-.03	+.44	+.68	-.52	+.24	-.15	+.30	+.09	-.07
G1 D2	-.10	+.75	+.69	-.38	+.06	+.08	-.05	-.09	.00
G1 D3	+.47	+.83	+.63	+.41	+.37	+.56	+.59	+.50	.00
G2 D1	-.13	+.61	+.53	+.04	-.04	+.12	-.28	-.32	.02
G2 D2	+.29	+.64	+.52	+.68	+.52	-.13	+.12	+.21	.00
G2 D3	+.42	+.66	+.52	+.07	+.44	+.43	+.34	+.17	.00

* N is generally 25-29. Data based on one 50-minute run vs. another. These data are for comparative purposes within this table. Negative coefficients can be taken as due to low reliability fluctuations.

TABLE 34

CORRELATION WITH RATINGS FOR INDEX OF ORDERLINESS

MEASURES, FACTOR SCORES, AND CONFLICTION MEASURES*

	Confliction Factor		Occupancy Factor		Comm. Factor		Delay Factor Score		Confliction Count (5 mi.)	
	SEM	CPM	SEM	CPM	SEM	CPM	SEM	CPM	SEM	CPM
G1 D1	-.23	-.22	+.11	+.04	+.21	+.08	-.36	-.37	-.23	-.21
G1 D2	-.16	-.08	-.01	+.06	-.23	-.19	-.25	-.26	-.29	-.17
G1 D3	-.24	-.25	+.18	+.01	-.34	-.22	-.58	-.52	-.33	-.24
G2 D1	-.48	-.38	+.05	-.03	+.15	-.05	-.23	-.25	-.35	-.23
G2 D2	-.35	-.31	-.16	-.21	+.32	+.17	-.20	-.21	-.28	-.24
G2 D3	-.35	-.24	+.20	+.26	+.13	+.03	-.37	-.43	-.28	-.19

	Conflicts Duration		ORD 1		ORD 1		ORD 3	
	SEM	CPM	SEM	CPM	SEM	CPM	SEM	CPM
G1 D1	-.29	-.24	+.22	+.24	+.02	+.06	+.04	+.04
G1 D2	-.17	-.07	-.23	-.11	-.15	+.01	-.03	+.05
G1 D3	+.04	-.03	+.11	+.08	-.03	+.10	.00	.00
G2 D1	-.20	-.19	+.11	-.04	-.01	-.06	+.07	-.09
G2 D2	-.15	-.19	-.26	-.23	-.34	-.29	+.03	-.12
G2 D3	-.14	.00	-.08	+.04	-.11	.00	.00	.00

* Data based on two-run aggregates; N is generally 27-31.

TABLE 35

MULTIPLE CORRELATION TO RATINGS WITH AND

WITHOUT INDEX OF ORDERLINESS MEASURES*

	R vs SEM	R vs CPM	N
Density One Sector 14 (D1 G1)			
Factors	.40	.38	31
Factors and "ORD 1"	.46	.51	31
Factors and "ORD 2"	.42	.43	31
Density Two Sector 14 (D2 G1)			
Factors	.34	.32	30
Factors and "ORD 1"	.43	.37	30
Factors and "ORD 2"	.35	.32	30
Density Three Sector 14 (D3 G1)			
Factors	.76	.64	29
Factors and "ORD 1"	.76	.66	29
Factors and "ORD 2"	.77	.64	29
Density One Sector 16 (D1 G2)			
Factors	.52	.47	31
Factors and "ORD 1"	.53	.48	31
Factors and "ORD 2"	.52	.47	31
Density Two Sector 16 (D2 G2)			
Factors	.50	.46	31
Factors and "ORD 1"	.50	.46	31
Factors and "ORD 2"	.50	.46	31
Density Three Sector 16 (D3 G2)			
Factors	.60	.60	30
Factors and "ORD 1"	.61	.60	30
Factors and "ORD 2"	.60	.60	30

* Data based on two-run aggregates.

TABLE 36

CORRELATIONS (r) BETWEEN TWO AVERAGED FACTOR

SCORES AND INDEX OF ORDERLINESS MEASURES*

		D1	D2	D3
ORD 1	G1	.67	.86	.88
	G2	.23	.37	.56
ORD 2	G1	.34	.71	.85
	G2	.18	.33	.63
ORD 3	G1	.18	.17	.00
	G2	.35	.07	.00

* Data based on two-run aggregates; N is generally 27-31.

RESPONSES TO POST-RUN QUESTIONNAIRES.

ANALYSIS. Questionnaires were given to the subjects of the two experiments in order to obtain their opinions on the realism of the simulation, any difficulties with the equipment, and their own opinion on the difficulty of the task and how well they were doing.

These data are of interest in that they provide an opportunity to examine the topics above, but also they provide an opportunity to examine some questions involving the relationships between these responses and other data in the experiment.

Similar questions were asked after each run in both experiments. The first question requested the controller to give a self-rating of the quality of the control technique which had been applied in the run just finished. The second question was meant to be an inquiry into system performance and was phrased as a question about the controllers' estimate of the feelings of the hypothetical pilots flying through the sector about how the system handled the traffic during the run. These two questions were on 7-point scales where the fourth box represented the average value. The third question asked for a comparison of the traffic level in the experimental run compared to the home sector. The fourth question asked about the realism of the simulator. These last two questions were on 5-point scales. When the data was coded for data reduction, numerical values were assigned to the rating scale positions. The questionnaires used in the two experiments, which were slightly different in phrasing although basically the same, are presented in Figures 2 to 5, in the discussion of procedures.

Tables 37 and 38 present the basic information about the questionnaire replies given by the average subject, for SEM I and SEM II, respectively.

In the SEM I experiment, the average controller thought technique was better in Geometry 2 than in Geometry 1, and better at lower densities than at higher denstities, although one should hasten to add that an interaction between sector and density is again apparent. A similar tendency is seen in the relative ratings given to what we have called above their rating of system performance. In these two items, the coding was such that a high number means the "good" end of the scale.

The SEM I question about traffic asked for a comparison between the traffic level in the simulation problem just completed and the difficulty in a peak hour at the home sector when serving as the radar controller having normal team support. Here, "much easier" was coded as a "1" in the data reduction and "much harder" here was coded as "5." Of course, the answers varied with sector and density. The difficulty of the highest SEM I traffic density was rated as somewhat higher than that they faced at home at peak hours, and the middle density as about the same, or slightly easier than, peak hour work with the assistance of the team. There was about a half's rating point difference between the two sectors in the middle density rating, indicating a slight

TABLE 37

MEAN VALUES OF QUESTIONNAIRE ITEM RESPONSES - SEM I

Item	Cell 1 G1 D1	Cell 2 G1 D2	Cell 3 G1 D3	Cell 4 G2 D1	Cell 5 G2 D2	Cell 6 G2 D3
1. Technique (1)	4.4	3.9	3.6	4.2	4.5	4.1
2. System (1)	4.5	3.9	3.4	4.4	4.5	3.8
3. Traffic Comparison (2)	1.7	2.8	3.4	1.7	2.4	3.3
4. Realism (2)	3.3	3.1	3.1	3.0	3.2	3.2

NOTES: (1) Rating scale 1 to 7
(2) Rating scale 1 to 5

TABLE 38

MEAN VALUES OF QUESTIONNAIRE ITEM RESPONSES - SEM II

Item	Day 1	Day 2	Day 3
1. Technique (1)	2.4	2.8	2.9
2. System (1)	3.3	3.8	3.9
3. Traffic Comparison (2)	2.8	3.0	3.0
4. Realism Comparison (2)	3.2	3.3	3.3

NOTES: (1) Rating scale 1 to 7
(2) Rating scale 1 to 5

feeling that geometry 2 was easier. Finally, in SEM I, the realism of the simulation process was considered adequate. In an open-ended question about the equipment, daily problems with the equipment were picked up and remedied. There were some complaints about the input devices on the radar consoles being different from those the controllers were used to in the field; this is now being remedied in a re-design of the simulator's controller positions.

For the SEM II experiment, the phrasing of three of the four rating questions was revised, although seeking similar information. In the first two questions, about the controller's own performance and the pilots' feelings about system performance, the wording was made more concrete, but the 7-point scales remained. Again, the poorer end of the scale was coded as "1" for the data reduction and the better end as "7." In responding to these first two items, the controllers generally regarded their performance in the runs about average for themselves, and felt that the system had performed at about an average level.

The rating item about the traffic was worded somewhat differently in the second experiment. The first experiment questionnaire had asked for a comparison of difficulty in the simulation hour exercise just completed with the difficulty in a peak hour in the home sector with the usual support; the second experiment items asked for a comparison of the traffic level just run to the traffic level which was usually encountered in the home sector, regardless of the team support used there. The direction of the scale and the coding were changed; a "1" in the second experiment's coding meant the traffic was considered heavier in the simulation and a "5" meant the traffic was heavier at home. Neither group of subjects expressed much difficulty with using these items.

On the first day, the SEM II traffic was rated somewhat heavier than the home sector traffic, where teams usually operate, as may be seen by the mean rating of 2.8 for day 1 in table 38. It will be remembered that this was approximately the same traffic level as had appeared in SEM I's geometry 1, density 2. There they had said it was about equal to the home sector's peak hour. On the second and third days, the traffic was rated at 3.0, or about the same as the traffic in the home sector.

In general, despite the differences in wording in the items, it can be said that they thought the traffic in these experiments was at least equal to the usual sector load in the field and somewhat higher and harder at times, as had been intentionally arranged, as was explained earlier under the topic of procedures and experimental design.

Turning now from the original purposes of the subject questionnaires of seeing how the subjects felt about the experimental runs as they proceeded, and of collecting information about equipment functioning, these data now might also be used to shed some light on some other questions of general interest.

In a general way, we might cónsider that there are four kinds of data here which might show interesting and informative relationships to one another,

TABLE 39

CORRELATIONS BETWEEN SUBJECT QUESTIONNAIRE ITEMS

AND OTHER DATA ITEMS

SEM I - CELL 1	Geometry (1), Density (1) Self Ratings			
	Technique	System	Traffic Comparison	Realism Comparison
Self Ratings				
Technique	1.00	0.37	0.09	0.29
System	0.37	1.00	0.06	0.74
Traffic Comparison	0.09	0.06	1.00	0.17
Realism Comparison	0.29	0.74	0.17	1.00
Observer Ratings				
SEM	0.24	0.38	-0.08	0.30
CPM	0.00	0.25	-0.12	0.17
Factors				
Confliction	-0.03	0.00	0.01	-0.05
Occupancy	0.31	0.44	0.32	0.38
Communications	-0.10	0.17	0.11	0.09
Delay	-0.05	-0.30	-0.09	-0.39
Measures				
N5C	0.03	-0.04	-0.22	-0.18
A/C Time Under Ctl.	0.36	0.46	0.32	0.40
Dur. G/A Contacts	-0.21	0.08	0.12	0.12
Total Delay Time	0.15	-0.16	-0.18	-0.21
# A/C Hdld	-0.04	0.08	-0.09	0.25
Fuel	0.36	0.42	0.31	0.32
N3C	-0.12	-0.09	0.22	0.05
# Delays	-0.08	-0.31	-0.07	-0.39

TABLE 39 (CONTINUED)

CORRELATIONS BETWEEN SUBJECT QUESTIONNAIRE ITEMS

AND OTHER DATA ITEMS

SEM I - CELL 2	Geometry (1), Density (2) Self-Ratings			
Items	Technique	System	Traffic Comparison	Realism Comparison
Self Ratings				
Technique	1.00	0.58	-0.19	0.47
System	0.58	1.00	0.26	0.79
Traffic Comparison	-0.19	0.26	1.00	0.35
Realism Comparison	0.47	0.79	0.35	1.00
Observer Ratings				
SEM	0.50	0.38	-0.13	0.41
CPM	0.48	0.38	-0.18	0.39
Factors				
Confliction	-0.27	0.12	0.21	0.04
Occupancy	-0.19	-0.04	0.24	0.06
Communications	-0.20	-0.04	0.17	-0.11
Delay	-0.38	-0.24	0.11	-0.13
Measures				
N5C	-0.29	0.06	0.12	-0.03
A/C Time Under Control	-0.20	-0.04	0.26	0.07
Duration G/A Contacts	-0.26	0.03	0.40	-0.08
Total Delay Time	-0.41	0.00	0.36	0.11
# A/C Hdld	0.04	-0.05	0.09	0.03
Fuel	-0.21	-0.06	0.25	0.02
N3C	-0.25	0.16	0.22	0.06
# Delays	-0.17	-0.36	-0.18	-0.29

TABLE 39 (CONTINUED)

CORRELATIONS BETWEEN SUBJECT QUESTIONNAIRE ITEMS AND OTHER DATA ITEMS

SEM I - CELL 3	Geometry (1), Density (3) Self Ratings			
	Technique	System	Traffic Comparison	Realism Comparison
Self Ratings				
Technique	1.00	0.51	0.06	0.02
System	0.51	1.00	0.16	0.43
Traffic Comparison	0.06	0.16	1.00	0.10
Realism Comparison	0.02	0.43	0.10	1.00
Observer Ratings				
SEM	0.56	0.63	-0.01	0.32
CPM	0.52	0.59	-0.11	0.28
Factors				
Confliction	-0.27	-0.10	0.24	0.10
Occupancy	0.01	0.03	0.20	0.13
Communications	-0.39	-0.47	-0.10	-0.05
Delay	-0.37	-0.40	0.21	-0.18
Measures				
N5C	-0.24	-0.03	0.11	0.02
A/C Time Under Ctl.	-0.02	0.01	0.18	0.18
Dur. G/A Contacts	-0.27	-0.38	0.30	-0.13
Total Delay Time	-0.35	-0.31	0.21	0.01
# A/C Hdld	0.29	0.24	0.08	-0.01
Fuel	-0.05	-0.01	0.14	0.07
N3C	-0.38	-0.20	0.16	0.06
# Delays	-0.36	-0.47	0.18	-0.39

TABLE 39 (CONTINUED)

CORRELATIONS BETWEEN SUBJECT QUESTIONNAIRE ITEMS
AND OTHER DATA ITEMS

SEM I - CELL 4 Geometry (2), Density (1)

	Self Ratings			
	Technique	System	Traffic Comparison	Realism Comparison
Self Ratings				
Technique	1.00	0.41	-0.38	0.11
System	0.41	1.00	-0.04	0.24
Traffic Comparison	-0.38	-0.04	1.00	0.01
Realism Comparison	0.11	0.24	0.01	1.00
Observer Ratings				
SEM	0.14	0.05	-0.36	-0.01
CPM	0.21	0.21	-0.44	0.12
Factors				
Confliction	-0.21	-0.11	0.28	-0.19
Occupancy	0.05	0.25	0.12	0.33
Communications	-0.23	-0.04	0.12	0.31
Delay	-0.05	0.03	0.10	-0.45
Measures				
N5C	-0.19	-0.12	0.23	-0.16
A/C Time Under Ctl.	0.02	0.23	0.07	0.31
Dur. G/A Contacts	-0.20	-0.10	0.04	0.29
Total Delay Time	0.16	0.27	0.12	-0.10
# A/C Hdld	-0.06	0.28	0.23	0.43
Fuel	-0.01	0.22	0.10	0.19
N3C	-0.11	-0.15	0.24	-0.14
# Delays	-0.22	-0.20	0.02	-0.53

TABLE 39 (CONTINUED)

CORRELATIONS BETWEEN SUBJECT QUESTIONNAIRE ITEMS
AND OTHER DATA ITEMS

SEM I - CELL 5	Geometry (2), Density (2) Self Ratings			
	Technique	System	Traffic Comparison	Realism Comparison
Self Ratings				
Technique	1.00	0.47	-0.09	0.10
System	0.47	1.00	0.01	0.42
Traffic Comparison	-0.09	0.01	1.00	-0.08
Realism Comparison	0.10	0.42	-0.08	1.00
Observer Ratings				
SEM	-0.19	-0.13	-0.12	-0.06
CPM	-0.10	-0.03	-0.37	0.06
Factors				
Confliction	-0.15	-0.06	0.20	-0.35
Occupancy	-0.11	-0.00	0.31	0.26
Communications	-0.09	0.11	0.24	0.18
Delay	-0.04	-0.03	-0.13	0.16
Measures				
N5C	-0.04	0.02	0.13	-0.33
A/C Time Under Ctl.	-0.14	-0.01	0.30	0.28
Dur. G/A Contacts	-0.07	-0.12	0.11	-0.04
Total Delay Time	-0.06	0.08	0.02	0.23
# A/C Hdld	0.33	0.11	0.02	-0.17
Fuel	-0.22	-0.01	0.32	0.26
N3C	-0.22	-0.18	0.20	-0.20
# Delays	-0.04	-0.23	-0.31	-0.09

TABLE 39 (CONTINUED)

CORRELATIONS BETWEEN SUBJECT QUESTIONNAIRE ITEMS

AND OTHER DATA ITEMS

SEM I - CELL 6

Geometry (2), Density (3)	Self Ratings Technique	Self Ratings System	Self Ratings Traffic Comparison	Self Ratings Realism Comparison
Self Ratings				
Technique	1.00	0.57	-0.16	0.07
System	0.57	1.00	0.05	0.25
Traffic Comparison	-0.16	0.05	1.00	-0.09
Realism Comparison	0.07	0.25	-0.09	1.00
Observer Ratings				
SEM	0.58	0.51	-0.20	0.20
CPM	0.56	0.46	-0.26	0.13
Factors				
Confliction	-0.43	-0.21	0.16	-0.49
Occupancy	0.07	0.25	0.09	-0.02
Communications	0.05	0.03	0.05	0.25
Delay	-0.12	-0.30	-0.21	-0.17
Measures				
N5C	-0.43	-0.24	0.19	-0.51
A/C Time Under Ctl.	0.02	0.18	0.09	-0.06
Dur. G/A Contacts	-0.06	-0.01	0.09	0.14
Total Delay Time	-0.09	-0.11	-0.05	-0.02
# A/C Hdld	0.12	0.24	0.17	0.12
Fuel	0.00	0.20	0.05	-0.05
N3C	-0.51	-0.23	0.18	-0.41
# Delays	-0.11	-0.36	-0.29	-0.24

omitting the rating on the simulation realism. The four kinds of data are:

a. Performance; Own opinion (subject)
b. Performance; Judge's opinion
c. Performance; Measured
d. Workload felt by subject (traffic level reply)

If this were merely a set of variables being intercorrelated, there would be ten possible inter-relationships here; but with four or more performance measures, depending on whether only the four factor scores or some others are used, there would be a considerably larger number of correlations. For this reason, the number of measures of each type will be restricted.

In SEM I, one such intercorrelation table was done for each cell (sector-density combination). In SEM II, one intercorrelation table was done for each day. The SEM II day data should be more informative since it is based on twice as many runs (four per day as compared to two per cell in SEM I). These tables appear as table 39 for the six SEM I sector-density cells and as table 40 for the three SEM II days.

Possibly the best way to approach this is by means of a series of single simple questions, all of which apply to both SEM I and II. Some questions of interest are:

a. What is the relationship between self-judged performance and other-judged (by observers) performance?

b. What is the relationship between self-judged performance and objectively measured performance?

c. What is the relationship between self-judged performance and self-judged workload?

d. What is the relationship between other-judged (by observers) performance and objectively measured performance?

e. What is the relationship between other-judged (by observers) performance and self-judged workload?

f. What is the relationship between self-judged workload and objectively measured performance?

Let us now examine these questions in an exploratory way, mainly to suggest hypotheses for other experimenters. The number of cases used for the correlations for the SEM I data is usually 29 to 31; and in the SEM II data, 39. The correlation value tabled as statistically significant (See appendix 3 for explanation) at the .05 level for 29 cases is approximately .37, for 39 cases is approximately .30.; only correlations above .30 will be looked at here.

The first question is: What is the relationship between self-judged performance and other-judged performance?

TABLE 40

CORRELATIONS BETWEEN SUBJECT QUESTIONNAIRE ITEMS AND OTHER DATA ITEMS

SEM II - DAY 1	Geometry (1), Density (2) Self Ratings			
	Technique	System	Traffic Comparison	Realism Comparison
Self Ratings				
Technique	1.00	0.44	0.18	0.35
System	0.44	1.00	0.35	0.10
Traffic Comparison	0.18	0.35	1.00	-0.12
Realism Comparison	0.35	0.10	-0.12	1.00
Observer Ratings				
SEM	0.20	0.47	0.29	-0.22
CPM	0.12	0.37	0.24	-0.23
Factors				
Confliction	-0.28	-0.10	-0.18	-0.11
Occupancy	-0.08	-0.07	-0.15	0.12
Communications	0.13	-0.24	0.00	0.24
Delay	-0.28	-0.53	-0.22	0.11
Measures				
N5C	-0.30	-0.22	-0.23	-0.12
A/C Time Under Ctl.	-0.08	-0.10	-0.16	0.10
Dur. G/A Contacts	0.06	-0.19	0.06	0.22
Total Delay Time	-0.25	-0.48	-0.23	0.14
# A/C Hdld	0.28	0.47	0.18	-0.14
Fuel	-0.11	-0.05	-0.20	0.02
N3C	-0.21	-0.10	-0.20	0.02
# Delays	-0.30	-0.58	-0.21	0.05

TABLE 40 (CONTINUED)

CORRELATIONS BETWEEN QUESTIONNAIRE ITEMS AND OTHER DATA ITEMS
SEM II

SEM II - DAY 2 — Geometry (1), Density (2)

	Self Ratings			
	Technique	System	Traffic Comparison	Realism Comparison
Self Ratings				
Technique	1.00	0.36	-0.05	0.48
System	0.36	1.00	0.16	0.34
Traffic Comparison	-0.05	0.16	1.00	-0.04
Realism Comparison	0.48	0.34	-0.04	1.00
Observer Ratings				
SEM	0.18	0.20	0.29	0.04
CPM	0.10	0.18	0.33	0.13
Factors				
Confliction	-0.13	-0.08	-0.12	-0.13
Occupancy	-0.23	-0.44	-0.43	0.19
Communications	0.01	-0.16	-0.29	0.16
Delay	0.07	-0.10	-0.24	0.35
Measures				
N5C	-0.09	-0.10	0.05	-0.08
A/C Time Under Ctl.	-0.16	-0.40	-0.45	-0.09
Dur. G/A Contacts	0.05	-0.11	-0.28	0.10
Total Delay Time	0.17	-0.02	-0.27	-0.25
# A/C Hdld	0.10	0.25	0.32	0.22
Fuel	-0.29	-0.39	-0.25	-0.35
N3C	-0.16	-0.15	-0.26	-0.15
# Delays	-0.04	-0.16	-0.17	-0.38

TABLE 40 (CONTINUED)

CORRELATIONS BETWEEN QUESTIONNAIRE ITEMS AND OTHER DATA ITEMS
SEM II

SEM II - DAY 3	Geometry (1), Density (2) Self Ratings			
	Technique	System	Traffic Comparison	Realism Comparison
Self Ratings				
Technique	1.00	0.42	-0.32	0.34
System	0.42	1.00	-0.10	0.56
Traffic Comparison	-0.32	-0.10	1.00	-0.02
Realism Comparison	0.34	0.56	-0.02	1.00
Observer Ratings				
SEM	-0.01	0.10	0.28	0.07
CPM	0.04	0.10	0.28	0.03
Factors				
Confliction	-0.25	-0.20	0.01	-0.23
Occupancy	0.21	-0.02	-0.44	-0.11
Communications	0.20	-0.05	-0.41	0.20
Delay	-0.02	-0.15	-0.11	-0.20
Measures				
N5C	-0.17	-0.26	0.04	-0.30
A/C Time Under Ctl.	0.24	-0.00	-0.45	-0.12
Dur. G/A Contacts	0.18	-0.04	-0.34	0.15
Total Delay Time	-0.06	-0.06	-0.05	-0.19
# A/C Hdld	-0.04	-0.30	-0.03	-0.16
Fuel	0.10	0.14	-0.38	-0.15
N3C	-0.13	-0.15	0.08	-0.14
# Delays	0.00	-0.17	-0.12	-0.19

If we consider the answer to this question to be obtainable from the relationship between the self-rating questions on technique and systems performance, on the one hand, and the observers' two ratings on the other, we can attempt an answer. The correlations between these subjective ratings by the observer and by the observed are sometimes encouraging, but fluctuate rather widely with the conditions, or are perhaps simply fluctuating on a sampling basis.

In the SEM I experiment, there is evidence of the expectable relationship, at least at the middle and high density levels, although somewhat more clearly in one sector rather than the other. In the middle density of Geometry 1, for example, there are correlations of .50 and .48 between the SEM and CPM ratings by the judges and the self-ratings of technique. Also, there are two positive correlations of .38 of the two observer ratings with the self-rating of system performance. Similar level correlations appear in two other cells, such as Geometry 1, Density 3, and Geometry 2, Density 3, and on Day 1 in SEM II for the system rating only.

The second question is: What is the relationship between self-judged performance and objectively measured performance?

Let us consider this question by examining the four factor scores and the self-ratings of technique and system performance. It is to be expected that these relationships will be negative, since a high self-rating should reflect a low number of conflictions, i.e., the scales run in opposite directions. In most cases, the correlations are indeed negative in sign. However, there are only a few correlations above .30. The primary and auxiliary raw score measures follow the factor scores in this, as usual.

The third question is: What is the relationship between self-judged performance and self-judged workload?

To answer this question, an examination was made of the correlation between the subject's rating of own technique and the rating of the traffic level faced. There are a few high correlations, but there seems to be no consistent pattern although there is a tendency to rate technique lower when the traffic seems heavier. It should be remembered that high number ratings in SEM I meant the subject felt the traffic was heavier in the simulation than at home, but in SEM II this scale was numerically reversed and a low coding number meant higher traffic.

The fourth question is: What is the relationship between other-judged (by observers) performance and objectively measured performance?

This one has already been answered at length in a previous section devoted to the subject. There it was found, at least when multiple correlations were used, that the relationships between objective scores and rated controller ability were substantial. Here, however, let us pause further over this question to simply illustrate a more graphic approach to the question of the relationhip between performance scores and controller ability, which might be examined further in the future.

Using SEM II day-level data, the controller judges' ratings for controller performance were arranged from lowest to highest. The four factor scores associated with those ratings were assembled into profiles for each individual, which was possible because they were on the same scale, as was discussed in the earlier discussion of both experiments' data having been put on the "third" scale. In figure 17, it could be said that it appears that the high performance controllers and the lower performance controllers may show different types of profiles. This constitutes a suggestion for further examination; much further work might be done in the realm of cluster analysis and profile analysis to explore such questions as the number of unique controller profiles of performance there might be.

The fifth question is: What is the relationship between other-judged (by observers) performance and self-judged workload?

In examining the SEM I correlations between the traffic question and the two obsevers' ratings, a few correlations in the negative thirties appear, -.36 and -.44 in the case of Geometry 2, Density 1, and -.37 in Geometry 2, Density 2. Apparently those who are functioning well in the opinion of the judges, at least, feel that the workload is lighter than others do. In the same data for the day level in SEM II, the correlations are close to thirty, but they are positive. This is probably a manifestation of the same phenomenon; the change in sign is understandable in that it may be remembered that the SEM II rating scale of traffic ran in the opposite direction from the SEM I rating scale.

The sixth question is similar to the fifth and is the following: What is the relationship between self-judged workload and objectively measured performance?

While there are not many correlations over .30 here, their directionality is appropriate. In SEM I if the controller felt that the traffic was heavy it would receive a higher numerical rating. In heavy traffic, most of the performance scores would naturally get higher (like delays). Therefore, positive correlations between the traffic ratings and the performance scores would be expected in the SEM I data, and this is generally the case. Because the SEM II scale on traffic ran in the opposite direction, essentially from "lighter here" coded as "1" to "heavier here" coded as "5," the SEM II correlations on this point would be expected to be opposite in sign and they usually are.

Finally, a word should be added here about an interesting relationship with the realism rating which was omitted from the earlier main discussion. There were some cases of positive correlations, some fairly high, between the subject's opinion of the realism of the simulation and the opinion held on the goodness of own-technique and system performance.

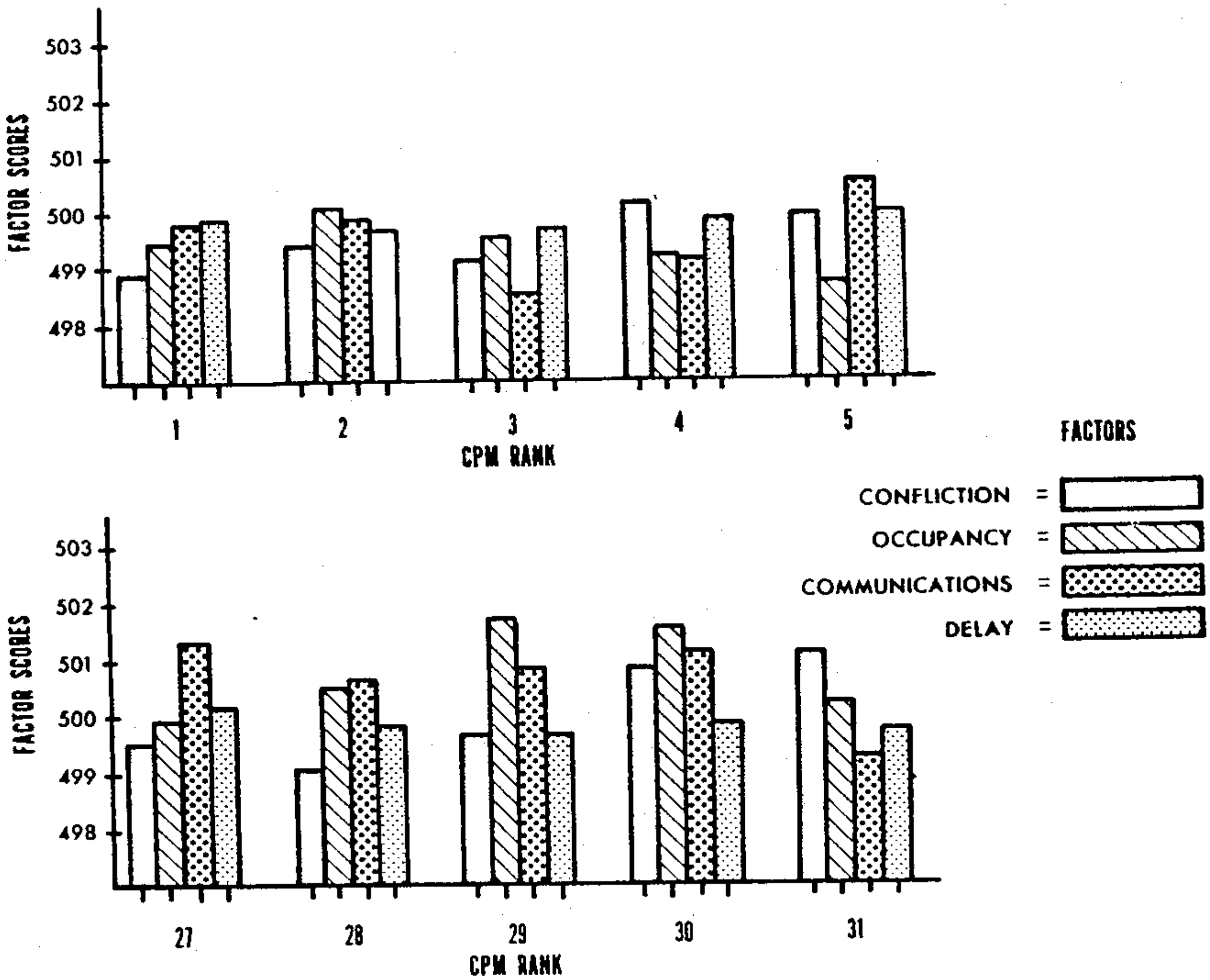

FIGURE 17. CONTROLLER PROFILES IN STANDARD SCORE FORM

IMPLICATIONS. The implications are:

1. The subjects felt that they did an average job, were not disturbed by any lack of realism, and felt that the traffic samples were tough; equal to "peak hour" with a full sector team helping them. The main purpose of this questionnaire, as has been said, was to check on daily experience, equipment functioning, and so on, and this purpose was fulfilled.

2. The data were adapted to make some explorations into the relationships between workload and performance, even though not ideally suited for the purpose. About all that can be said here is that such relationships, if they exist, are weak and situation-dependent.

DISCUSSION

There is no question as to whether real-time air traffic control system simulation will be used in the future. It seems an eminently worthwhile, albeit expensive, thing to do. Although people feel that they get information out of it about air traffic control system problems and issues, the real question is whether they get information or misinformation.

Here enters a true philosophical issue. Are impressions information? If someone watches a controller use a proposed system, and thinks it functions better than the current system in use, is that information? If the controller is asked for an opinion and gives it, is that information? Suppose that the traffic mix or level or procedures are somewhat different from those which the controller or the observer are used to. Are their impressions dependable enough to base huge expenditures for new systems on them? Suppose the designer of the new system is giving his observation, is that information?

These are the kinds of considerations that make objective measurement and statistical techniques desirable. It is because grave errors can be caused by subjectivity in interpreting what is seen, and sometimes even in interpreting what has been genuinely measured, as, for example, when the hypothesis about what measures shall be considered important has not been stated in advance. However, measurement of the joint performance of human and machinery in accomplishing the mission of an information processing and decision making system is not a simple task. To develop methods and measures for such a purpose is a difficult, time consuming and risky effort. It must be remembered that the performance under study is not rote or mechanical but very dynamic. The thing to be surprised about is not that the measurement process may be discouraging, but that there is anything encouraging about it at all.

There may, in fact, be a middle ground possible between sheer impressionism and strict empiricism. This might consist of carefully controlled and administered observation and rating forms being given to trained, impartial and fresh observers. But even this would be in need of an evaluation and refinement process.

The worst case of all, though, is the one that appears to be more frequent and customary than even those who engage in it acknowledge. Simply stated, these are studies in which the investigators, in all good faith, use objective measurements that can be obtained from a simulator apparently without realizing that such measures, even though numerical, are behavior and performance measures and have a wide band of error around them.

On the other hand, only the most crucial system evaluations, perhaps, need to be conducted using strict inferential rules. There are times, as Stammers and Bird (reference 14) say, using the Sinaiko and Belden term (reference 15), when the proper thing to do is the "indelicate experiment." The work by Stammers and Bird concerned a data transfer and display system for airport controllers and was carried out for the Royal Radar Establishment. It is a fine example of such an effort. Another type of brief and uncomplicated simulation being a good idea is when it is done for the purpose of exploring concepts as part of a long continuing examination. What appears to be an example of this is the work of Tobias and O'Brien on RNAV (area navigation) for NASA (reference 16).

In working on evaluating human factors aspects of computer aiding for air traffic controllers, Whitfield, Ball and Ord (reference 17) achieved a good integration of the best features of the "indelicate" experiment and the more traditional experiment.

The topics of methods and measurements in the air traffic control system have been discussed at length by Hopkin (references 18 and 19) and the general topic of systems experimentation involving performance measurement has been discussed in a book by Parsons (reference 20).

While admitting that various degrees of indelicacy may be permissible depending on the circumstances, it is still important to pursue the ideals of classic experimentation where possible and appropriate. That being the case, let us review some of the "lessons learned", which might be of use in pursuing both the delicate and indelicate experiment.

The first and most important lesson was also and first pointed out by Horowitz (reference 13), and it is to consider the beta error. As Horowitz pointed out, people in medicine and medical research do this all the time and people in other practical fields should do so too. What he had encountered was the tendency in some statistically minded people to set the level of the alpha error they will accept at the traditional .05 level, and to ignore the beta error. Especially with difficult data such as is found in dynamic simulations, this leads to frequent, if not continual, failure to reject the null hypothesis. In a practical sense, that sort of uncritical application of statistical techniques could lead to the rejection of many fine system concepts. This is what Horowitz rightly pointed out.

The data from these experiments and the power tables based on them can reduce the likelihood of that kind of error by asking that the levels of alpha and beta that will be used and the amount of difference it is sought to detect be specified in advance. It is possible to compensate for the lack of statistical robustness in the measurement process by choosing moderate levels of these parameters.

A second major lesson learned is the importance of the practice and familiarization factors in system experiments and evaluations. The learning curves sought and found in the SEM II experiment were quite dramatic. For this reason, careful thought must be given to practice effects and hence sequence effects in the design of such experiments. However, it should be of some assistance to know how long these effects last, as indicated by the curves.

Related to the question of the statistical power of simulation data, is the question of the reliability (repeatability) of such data. While one can compensate for such unreliability as was found here, it was found to be lower than expected based on the only other experiment having such data. It was, in fact, expected to be some amount higher since now the data was being collected by computer instead of by paper and pencil. This did not turn out to be the case. While the reliability was not totally discouraging, due to the fact that it can be compensated for by means of considering the setting of the alpha and beta levels actually needed and by data aggregation, it is puzzling. This is a topic which deserves some careful work and thought. The lesson to be learned here is that new ideas for system measures should be sought out on a continuing basis.

This same unreliability should caution those who wish to run simulations to the effect that if a single sector system is comparatively unreliable, then a multi-sector simulation's data are almost certainly much more unreliable, because of the additional sources of variance introduced. While the reliability and power calculations made here do not apply to multi-sector simulations, they can be regarded as an optimistic estimate of what would occur in a larger simulation.

While on the subject of the single-sector, single-controller system, it did, of course, include simulated conversation and coordination with adjacent sectors and even terminal areas, and, while we obtained no evidence on this topic beyond the subject controllers' ratings of simulation realism, it seemed quite satisfactory as a method for simulating the essence of the controller's job. It would seem to recommend itself as a rather economical way of studying many man/machine interface problems or plans, and even as a way to evaluate individual controller training progress.

Another lesson learned here was that we only need to analyze a comparatively small number of measures: the four factor scores, the four primary measures, and the two additional auxiliary scores. This makes an enormous difference in the sheer feasibility of data handling chores and interpreting this kind of data. This set of scores should be accepted as an operating base for all enroute simulations, at least until something better comes along, and programmed into the simulation data collection system. A bonus from this practice would be that after some time all ATC system simulations would be interpretable in common data distribution terms.

Excessive reliance on ratings by judges is not recommended even though the judges here performed with some reliability. It must be remembered that they were carefully and deeply trained, and were constantly observing the same exercises.

Another lesson which should be learned is that there is available a way to accumulate a set of traffic problems which are extremely different, thus reducing practice effects, but which can be shown to be of a comparable level of difficulty. The interaction between sector geometry and traffic density could be used to generate a library of traffic samples whose level of difficulty, as indicated by score distributions obtained in small experiments, could be considered interchangeable. Another way of handling the traffic sample "same but different" requirement was also demonstrated here, the shuffling of start times in the same level-profile trafffic sample.

The main lesson to be learned from the experience with the index of orderliness was not a clear-cut lesson about that index, which did not emerge, but, nonetheless, a demonstration that, given a data base like that used here, many investigations about different and novel measures might be conducted.

A major question which arises is that of whether there is additional work in this area which should be done. There are at least three study efforts which should be undertaken, and it should be pointed out at the outset that accomplishing them will be considerably easier because of that which has been done so far since, in subsequent investigations, even of methodology, the power estimates which are available will enable careful planning of the required size of the experiments which are to be conducted.

The first and most obvious follow-on work would involve continuing to work with the available data bases from these experiments in order to seek for refined measures. It should be remembered that the focus in this effort so far was evaluative, not developmental. As a next step, various ideas for novel measures could be computed in these data bases and their relationships to one another and to the standard measures already present could be examined.

The second step would be to extend the method to a multiple-controller sector team and to a multi-sector system of reasonable size, say three sectors. The goal here would be the comparatively simple one of determining the change in variance, power and reliability which would be caused by working with these more complex system spaces. This would probably be desirable to do even though, on the one hand, it would be hoped that the need people feel for duplicating complex system spaces in simulation would be diminishing, and, on the other hand, that the present power estimates could be used as approximations (albeit optimistic ones for large systems).

The third possible direction would be to make a start into the study of terminal area simulation methodology and measurements. A beginning on this had been made, but has since been postponed. In basic outline, the approach that had been tentatively decided upon was as follows. First, there was to be an assembly of the customary classic measures for terminal area air traffic control system functioning. Next, these measures would be administered at three levels of traffic density and with several replications to a large number of control "teams." The first attempt would be to try to reduce the number of measures by searching for the basic dimensions of measurement, and having found those, to examine the data to estimate the parameters needed to plan experiments of desired levels of statistical power for system evaluations.

However, the terminal area air traffic control system is nowhere near as simple as the en route system. It is easy and clearly legitimate to represent the en route system in microcosm; but the terminal system does not readily lend itself to such simplification. The terminal team is composed of several individuals working not on the same airspace but on different parts of the airspace. While the smallest en route team groups around one radar picture, the smallest terminal team might consist of an arrival controller, a departure controller, and a local controller and ground controller. While ground and local control have rarely been simulated, they could be by use of some simplifying assumptions and rough presentations. Specifically, it would probably not be too unrealistic to use the simulator to show the airport surface as if on radar for the purpose of running a complete simulation. Doing such a simulation was considered. Also considered was running, with the same people, a terminal area simulation in which one controller was looking at the entire terminal area and performing the total control function alone, at a much reduced level of traffic, of course. One major purpose would be to determine if the same measures were statistically important in both team and microcosm (single controller) terminal simulations. Another purpose would be to determine if, when similar conditions (systems, geometries, etc.) were compared in team and microcosm simulations, similar outcomes resulted. This would render many terminal area issues investigatable by simulation which are now almost prohibitive in the amount of effort required to accomplish them. Progress was made in developing the list of measures which was to be evaluated and it is presented as Appendix F for the use of those who might be engaged in terminal area simulation work.

There is one last comment it seems important to make about possible future research that this experience has suggested. This, briefly, has to do with the application of the methodology developed here to a related field, as a training progress criterion measure device for the individual controller. While, as said earlier, the reliability needs considerable improvement for such a purpose, such improvement does not seem impossible.

CONCLUSIONS

These experiments provided a statistical and methodological baseline for quantitative system assessment using real-time air traffic control simulation testing. In particular, the following conclusions have been reached:

1. The en route measure set as presently constituted forms recognizable operationally meaningful clusters of measures. These are confliction, occupancy, communication and delay.

2. The four factor measures produce as valid an assessment of system performance as do the original many raw measures.

3. The acquisition of stable data requires six hours of preliminary familiarization and training in the experimental environment.

4. The same four factors were tried in another experiment with another sector geometry, two additional traffic densities, and a different group of controllers, the factors still held up as being adequate basic dimensions of measurement.

5. System evaluation using real-time ATC simulation in an objective manner is only possible in a technically sound way if account is taken in planning experiments of the relatively low statistical power of the measurement which can be accomplished in the dynamic exercises. Tables of the statistical power of the basic factor scores have been assembled based on the data here collected and analyzed. Failure to assure adequate power will in most system evaluations lead to the rejection of actually promising system ideas.

It is to be emphasized that the above conclusions were reached during tests where one person, serving as the radar controller for the sector, was responsible for all the traffic in the sector. Also, the traffic density was held at a relatively constant level throughout a given session. However, adequate provision for the exercise of adjacent sector coordination was included, and some of the assistant controller duties were pre-performed. It seems certain that the "one-person team" procedure would not have affected the basic dimensions of measurement found for system effectiveness; although the estimates of inter-team variation which entered into the power calculations might possibly have been affected.

REFERENCES

1. Bona, L. J. and Pszczolkowski, S. R. The FAA's Air Traffic Control Simulation Facility. in: Technology in air traffic control training and simulation., Proceedings, Vol. II, Second International Learning Technology Congress and Exposition on Applied Learning Technology., February 14-16, 1978. Society for Applied Learning Technology, Warrenton, Virginia.

2. Buckley, E.P., O'Connor, W. F. and Beebe, T. A Comparative Analysis of Individual and System Performance Indices for the Air Traffic Control System, Federal Aviation Administration, National Aviation Facilities Experimental Center, Report NA-69-40, 1969.

3. Dixon, W. J. Biomedical Computer Programs. University of California at Los Angeles Press, Berkeley, California, 1975.

4. Harman, H. H. Modern Factor Analysis. 2d edition. University of Chicago Press, Chicago, Illinois. 1979.

5. Boone, J. O. et al. The Federal Aviation Administration's Radar Training Facility and Employee Selection and Training. FAA-AM-80-15. CAMI, FAA, Oklahoma City, Okla., Sept. 1980.

6. Boone, J. O. and Steen, J. A. A Comparison Between Over the Shoulder and Computer-Derived Measurement Procedures in Assessing Student Performance in Radar Air Traffic Control. Aviation, Space and Environmental Medicine, 52(10): 589-593, 1981

7. Cobb, B.B. The Relationships Between Chronological Age, Length of Experience, and Job Performance Ratings of Air Route Traffic Control Specialists. FAA-CAMI Report AM 67-1, June, 1967.

8. Gaebelein, J.W. and Soderquist, D.R. The Utility of Within-Subject Variables Estimates of Strength. Educational and Psychological Measurement, 1978, 38, p. 351-9.

9. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York. Academic Press. 1969.

10. Halvorsen, A. G. Definition and Description of System Performance Measure Entitled Index of Orderliness for Use in Digital ATC Simulation Studies. National Aviation Facilities Experimental Center: Working Paper, September, 1970.

11. Gent, H. A Measuring Rod for ATC Systems: The Index of Orderliness. ATC Symposium, Boston, MA, 1975.

12. Mengert, P. H. and Englander, T. Preliminary Statistical Validation Index of Orderliness ARTS III Test Data. U. S. D.O.T. Transportation Systems Center: Final Report, April 19,1971.

13. Horowitz, Seymour M. An Evaluation of the ARTS III Level of Automation.(third lot procurement). July, 1972. FAA Office of Aviation Economics.Final Report.
Not numbered.
14. Stammers, R. B. and Bird, J. M. Controller Evaluation of a Touch Input Air Traffic Data System: An "Indelicate" Experiment. Human Factors, 1980, 22(5), 581-589.

15. Sinaiko, H. W. and Belden, T. G. The Indelicate Experiment. in: J. Speigel and D. E. Walker (Eds.), Second Congress on the Information System Sciences. Washington, D. C.: Spartan Books, 1965, 343-348.

16. Tobias, L. and O'Brien, P. Real-Time Manned Simulation of Advanced Terminal Area Guidance Concepts for Short-Haul Operations. NASA, Washington DC, August 1977. NASA Technical Note NASA TN D-8499.

17. Whitfield, D., Ball, R. and Ord, G. Some Human Factors Aspects of Computer-Aiding Concepts for Air Traffic Controllers. Human Factors, 1980,22(5), 569-580.

18. Hopkin, V. D. Conflicting Criteria in Evaluating Traffic Control Systems. Ergonomics, 1971, 1971, 14,557-564

19. Hopkin, V. D. The Measurement of the Air Traffic Controller. Human Factors, 22(5), pp.547 to 560.

20. Parsons, H. M. Man-Machine System Experiments. Baltimore: The Johns Hopkins Press, 1972.

21. Mc Nemar, Q. Psychological Statistics. 3rd edition. New York, Wiley, 1962.

22. Winer, B. J. Statistical Procedures in Experimental Design. 2d edition. New York, Wiley, 1971.

23. Kirk, R. E. Experimental Design: Procedures for the Behavioral Sciences. Brooks/Cole, Belmont, CA. 1968.

24. Guilford, J. P., Fundamental Statistics in Psychology and Education. Fourth Edition. New York, Wiley, 1965.

APPENDIX A

LIST OF SYSTEM EFFECTIVENESS MEASUREMENTS AND DEFINITIONS:

SEM EXPERIMENT I

1. TGT Spacing Analysis (A)

   A count of the number of instances two aircraft violate the separation allowance of 950 feet vertically and 4 miles horizontally.

2. TGT Spacing Analysis (B)

   Same as above with 5 mile horizontal separation allowance.

3. TGT Spacing Analysis (C)

   Same as above with 3 mile horizontal separation allowance.

4. Number of Start Delays

   A count of the number of instances an aircraft entered the system at a time greater than its scheduled time (plus two minutes).

5. Start Delay Time

   The duration of the start delays (Measure 4).

6. Number of Hold and Turn Delays

   A count of the number of holding delays plus a count of the number of turn delays lasting more than 100 seconds.

7. Hold and Turn Delay Time

   The duration of the hold and turn delays (Measure 6).

8. Number of Arrival Delays

   A count of those start delays of arriving aircraft.

9. Arrival Delay Time

   The duration of arrival delays.

10. Number of Departure Delays

A count of those start delays of departing aircraft.

11. Departure Delay Time

The duration of departure delays.

12. Time in System

The number of active aircraft controlled by the subject, incremented each second that control was exercised.

13. Number Aircraft Handled

Total number of aircraft under subject's control.

14. Number of Completed Flights

The number of flights terminated by a handoff.

15. Number of Arrivals Achieved

A count of enroute traffic transferred to the termination frequency.

16. Number of Departures Achieved

A count of active departures.

17. Arrival Altitudes not Attained

A count of enroute arrivals not transferred to the termination frequency at an altitude greater than was predetermined, plus 100 feet.

18. Departure Altitudes not Attained

A count of enroute departures not transferred to the termination controller at an altitude less than was predetermined minus 100 feet.

19. Number of Contacts

A count of ground to air microphone contacts.

20. Communication Time

The duration of ground to air contacts (Measure 19).

21. Number of Altitude Changes

A count of pilot messages to alter aircraft altitude.

22. Number of Heading Changes

A count of pilot messages to change heading.

23. Number of Speed Changes

A count of pilot messages to revise aircraft speed.

24. Number of Handoffs

The number of acknowledged handoffs to the subject.

25. Handoff Delay Time

The time between a handoff and the subject's acceptance of that aircraft.

26. Re-idents

A count of beacon identity requests.

APPENDIX B

LIST OF SYSTEM EFFECTIVENESS MEASUREMENTS AND DEFINITIONS: SEM EXPERIMENT II

Given below is a list of measures used in this experiment with definitions and commentary. They generally consist of event counters with their respective duration. All duration measures are in seconds.

Unless noted to the contrary, all measures are keyed to the following rule to determine if an aircraft is under the control of the subject.

CONTROL RULE

An aircraft is under control if it is within the sector boundary or on the frequency of the subject.

That is to say, in order for an aircraft not to be under control it must be both outside the sector and off the subject's frequency. When under control, an aircraft is considered the subject's responsibility and all events relative to that aircraft are charged to the subject.

DA - 01 Number of Path Changes (PTHCHAD)

The number of altitude, heading, and speed change messages sent to aircraft under control.

DA - 02 Number of Barrier Delays (BRNDELD)

The number of instances a subject asks that all entering traffic be halted.

DA - 03 Duration of Barrier Delays (BRDURAD)

The cumulative time that barrier delays remain in effect. The beginning of a barrier delay is referred to as a STOP message and its termination as a START message.

DA - 04 Number of Start Delays to Aircraft (NSTADLD)

The number of instances that an aircraft was scheduled to enter the problem while a STOP message was in effect.

Note that STOP and START messages can occur without any start delays accumulating. (If, for instance, no aircraft were scheduled to enter during this interval.)

DA - 05 Duration of Start Delays to Aircraft (DSTADLD)
The cumulative duration of start delays. For each affected aircraft, the start delay equals the difference between its scheduled start time and the time a START message is entered.

When traffic is stopped and then restarted all aircraft have their problem entry time adjusted to keep the original spacing intact.

DA - 06 Number of Hold and Turn Delays to Aircraft (NOHTDLD)
The number of occasions that aircraft are put into a hold or a turn lasting more than 100 seconds. This is counted aircraft under control.

DA - 07 Duration of Hold and Turn Delays to Aircraft (TMHTDLD)
The cumulative time of hold and turn delays.

Note that hold and turn delays occur only within the sector, and that turn delays are counted only after 100 seconds. This is to allow course changes to be counted as such.

DA - 08 Number of Handoffs Accepted (NOHDFAD)
The number of aircraft handed off and accepted by the subject controller.

DA - 09 Handoff Acceptance Delay Time (HDFDELD)
The cumulative time between a handoff and the acceptance of that aircraft by the subject.

DA - 10 Number of Contacts (Ground to Air) (NOCTCSD)
The number of times microphone transmission is made by the subject.

DA - 11 Duration of Contacts (Ground to Air) (DUCTCSD)

The cumulative time of ground to air contacts.

DA - 12 Total Delays (Hold + Turn + Start) (TODLYND)

(DA - 04) + (DA - 06)

DA - 13 Total Delay Time (TODLTID)

(DA - 05) + (DA - 07)

DB - 01 Number of Aircraft Handled (NACHDLD)

The number of aircraft that are accepted onto the subject frequency, or enter the sector. (See Control Rule above.)

DB - 02 Aircraft Time Under Control (ACTUC D)

The amount of time aircraft are under control, summed for all aircraft handled.

DB - 03 Average Aircraft Time Under Control (ACTUCAD)

(DB - 02) divided by (DB - 01)

DB - 04 Target Spacing Analysis - A (TSA A4D)

The number of instances that aircraft violate the separation standard of 4 miles horizontal spacing and 950 feet vertical spacing. At least one of the aircraft involved must be under control (see Control Rule above).

The measure is also referred to as 4 mile conflicts.

DB - 05 Target Spacing Analysis - B (TSA B5D)

Same as above with 5 mile horizontal separation.

DB - 06 Target Spacing Analysis - C (TSA C3D)

Same as above with 3 mile horizontal separation.

DB - 07 Duration of TSA-A (DURTSAD)

The cumulative duration of 4 mile conflicts.

DB - 08 Duration of TSA-B (DURTSBD)

The cumulative duration of 5 mile conflicts.

DB - 09 Duration of TSA-C (DURTSCD)

The cumulative duration of 3 mile conflicts.

DB - 10 Aircraft Distance Flown (ACDSTFD)

The cumulative distance in miles flown by aircraft while under control.

DB - 11 Fuel Consumption (FUELCOD)

The cumulative fuel in pounds consumed by aircraft under control.

DB - 12 Number of Completed Flights (NCPFTSD)

The number of aircraft accepted by the subject that reach their destination and are transferred by frequency change. Control, as defined by the Control Rule, must be relinquished at the destination point to be counted as a completed flight.

Note that flights under control when the data period begins are completable.

DB - 13 Arrival Altitudes Attained (ARVLATD)

The number of arrival aircraft whose flight is completed within 100 feet of their goal altitude.

DB - 14 Departure Altitudes Attained (DPTRATD)

Same as above for departure aircraft.

DB - 15 Aircraft Time in Boundary (ACBTM D)

The cumulative time that aircraft under control are within the test sector.

SEM System Effectiveness Measure

(See Appendix C)

CPM Controller Performance Measure

(See Appendix C)

APPENDIX C

DEFINITIONS AND USAGES

DEFINITIONS AND USAGES

A. Definitions: Defined here are technical terms from the area of statistics. An attempt has been made here to define the terms in a non-technical manner. Also given are sourcebooks, usually textbooks, where more detailed definitions can be gleaned.

TERM	SOURCEBOOK	DEFINITIONS (not necessarily direct quotes from Sourcebook)
Experimental design	Kirk, Winer	A plan for the collection of data which includes the form of analysis which will be applied including the hypothesis to be tested.
Statistical significance	Mc Nemar	An outcome of a test is said to be statistically significant when the calculations and tests indicate that the probability is small (of a certain pre-determined small size) that such an outcome could have occurred by chance.
Coefficient of correlation	Mc Nemar	The coefficient of correlation expresses the degree of linear relationship between two variables. It ranges from - 1.00 (an inverse perfect relationship) to zero relationship through + 1.00 (positive perfect relationship). Signified by r.
Standard error of measurement	Mc Nemar	If it were possible to take a large number of repetitions of the same run in the exact same circumstances, their scores would normally distribute themselves around the "true" value. Using this, it is possible to say the chances that the "true" value is within certain bounds around the obtained value.

TERM	SOURCEBOOK	DEFINITIONS
Intra-class correlation	Mc Nemar	A correlation form used in cases where there is no prior reason to assign a score to one of the two distributions being correlated; in this case, the members of the pairs of judges.
Multiple correlation coefficient	Mc Nemar	A correlation between a linear combination of variables on the one hand and another single variable on the other hand. Signified by R.
Logarithmic transformation	Winer	Taking the logarithm of each score in a distribution of scores sometimes results in a distribution more closely resembling the normal distribution.
Coefficient of determination	Mc Nemar	The square of the correlation coefficient (simple or multiple) expresses the common variance between the two variables, is the variance in one accountable for by variance in the other.
Factor analysis	Guilford	A statistical technique which uses the correlation among measures to find the minimum set of measures which adequately expresses the same information in a more condensed manner.
Correction for shrinkage	Mc Nemar	In multiple correlation, this formula can be applied to correct R for the number of predictors. If the number of predictors (n) approaches N (the number of cases) there is a bias. The formula corrects for that bias and makes allowance for a decrement in R which occurs when applied to new samples of subjects. The formula is $R'_{1.23..n} = \sqrt{1-(1-R^2)\left(\frac{N-1}{N-M}\right)}$

TERM	SOURCEBOOK	DEFINITIONS
Phi coefficient	Harman	A coefficient expressing the relationship or resemblance between factor analysis weights from two sources.
Statistical significance, r, R	Guilford	A size of correlation coefficient (simple or multiple) which is large enough to be sure with a probability of 5/100 chances of error that there is a relationship between the variables (or combinations of variables).
Analysis of Variance	Winer	A statistical technique used to test statistical inference hypotheses. Basically it compares the variance in scores across conditions (systems) with the natural variance among peoples' performances.
Standard scores	Mc Nemar	A standard score distribution is created by expressing each score in a distribution as a deviation from the mean of the original distribution and dividing this difference by the standard deviation of the distribution.

B. <u>Usages:</u> Defined or explained here are words having particular usages in this report.

Day-level - In the SEM II experiment, four exercises were usually run in a day. Thus, a four-run aggregate is usually referred to as a "day-level" figure.

Traffic sample - A group of aircraft with flight plans which are scheduled to enter the simulated air traffic control system at designated entry times.

System Effectiveness (SEM) Rating - A rating of the effectiveness of the simulated air traffic control system under study. These were made by the controller observer-judges. By effectiveness is meant here the degree to which the system achieves its missing of the safe and expeditious movement of aircraft.

Controller Performance (CPM) Rating - A rating of the skill and technique of the subject controller performing the air traffic control exercise.

Measure - An objective quantitative recording of some aspect of a phenomenon. The basic measures used here were believed to be indicators of the quality of air traffic control such as the number of aircraft handled, the number of aircraft delayed, etc.

Factor score - A composite score based on several measures as was indicated by the factor analysis of the initial set of measures.

Subject - An air traffic control specialist who controls traffic in the simulated ATC system. To be distinguished from a "ghost" or support controller who plays the role of adjacent air traffic control facilities.

Primary score - One of a set of four of the original measures which was chosen as capable of being a good representative of one of the four factors.

Auxiliary score - Two of the original set of measures which were not picked out by the factor analysis, but it was decided were important to keep in the final measure set; these were the number of aircraft handled and full consumption.

Common scale - A common scale is one which is in the same units (such as is the case with such units as feet, pounds, etc., as examples). While not so readily visualizable, standard score scales adjusted to have the same mean and standard deviation have the same desirable quality (see standard scores under "Definitions."

APPENDIX D

SUPPLEMENTARY TABLES

TABLE 1. SORTED ROTATED FACTOR LOADINGS (PATTERN), DAY 1

	Conflict Factor	Occupancy Factor	Communication Factor	Delay Factor
Duration Target Spacing Analysis A	.926	*	*	*
Duration Target Spacing Analysis C	.864	*	*	*
Duration Target Spacing Analysis B	.854	*	*	*
Target Spacing Analysis A	.836	*	*	*
Target Spacing Analysis C	.824	*	*	*
Target Spacing Analysis B	.708	*	*	*
Time Under Control	*	.968	*	*
Fuel Consumption Under Control	*	.950	*	*
Aircraft Distance Flown Under Control	*	.924	*	*
Aircraft Time in Boundary	.311	.588	*	*
Handoff Accept Delay Time	*	.583	*	.515
Number of Ground to Air Contacts	*	*	.887	*
Duration of Ground to Air Contacts	*	*	.874	*
Path Changes	*	*	.791	*
Total Delay Time	*	*	*	.926
Total Delays	*	*	*	.924
Arrival Altitude Attained Completed Flights	*	*	*	*
Variance Accounted For	4.464	3.531	2.406	2.306

The above factor loading matrix has been rearranged so that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than .5000 appear first. Loadings less than .2500 have been replaced by *.

TABLE 2. SORTED ROTATED FACTOR LOADINGS (PATTERN), DAY 2

	Conflict Factor	Communication Factor	Occupancy Factor	Delay Factor
Target Spacing Analysis A	.935	*	*	*
Target Spacing Analysis B	.892	*	*	*
Duration Target Spacing Analysis B	.859	*	*	*
Duration Target Spacing Analysis A	.834	*	*	.313
Target Spacing Analysis C	.724	*	*	.337
Duration Target Spacing Analysis C	.630	*	*	.620
Arrival Altitude Attained Completed Flights	.508	.363	*	*
Number of Ground to Air Contacts	*	.898	*	*
Duration of Ground to Air Contacts	*	.868	.284	*
Path Changes	*	.861	*	*
Time Under Control	*	*	.851	*
Time in Boundary	*	.336	.850	*
Fuel Consumption Under Control	*	*	.665	*
Total Delay Time	*	*	*	.869
Total Delays	*	*	*	.856
Aircraft Distance Flown Under Control	*	*	.346	*
Handoff Accept Delay Time	*	*	*	*
Variance Accounted For	4.370	2.657	2.305	2.174

The above factor loading matrix has been rearranged so that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than .5000 appear first. Loadings less than .2500 have been replaced by *.

TABLE 3. SORTED ROTATED FACTOR LOADINGS (PATTERN), DAY 3

	Conflict Factor	Occupancy Factor	Communication Factor	Delay Factor
Duration Target Spacing Analysis A	.853	*	*	*
Target Spacing Analysis C	.847	*	*	*
Target Spacing Analysis A	.845	*	*	*
Target Spacing Analysis B	.829	*	*	*
Duration of Target Spacing Analysis C	.819	*	*	*
Duration Target Spacing Analysis B	.701	*	*	*
Time Under Control	*	.964	*	*
Time in Boundary	*	.865	*	*
Fuel Consumption Under Control	*	.853	*	*
Aircraft Distance Flown Under Control	*	.677	*	*
Duration of Ground to Air Contacts	*	*	.906	*
Number of Ground to Air Contacts	*	*	.885	*
Path Changes	*	*	.851	*
Total Delay Time	*	*	*	.868
Total Delays	*	.385	*	.828
Arrival Altitudes Attained Completed Flights	*	*	*	*
Handoff Accept Delay Time	*	*	*	*
Variance Accounted For	4.097	3.134	2.403	1.639

The above factor loading matrix has been rearranged so that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than .5000 appear first. Loadings less than .2500 have been replaced by *.

TABLE 4. CONFLICT FACTOR

Factor Score Coefficients

	Day 1	Day 2	Day 3
Path Changes	-.02	.03	-.03
Handoff Accept Delay Time	.03	-.02	.00
Number of Ground to Air Contacts	.03	.01	-.01
Duration of Ground to Air Contacts	.04	.00	.01
Total Delays	-.06	-.07	-.03
Total Delay Time	-.07	-.04	-.03
Time Under Control	.01	.01	.01
Target Spacing Analysis A	.17	.23	.19
Target Spacing Analysis B	.10	.24	.18
Target Spacing Analysis C	.19	.15	.26
Duration Target Spacing Analysis A	.25	.17	.20
Duration Target Spacing Analysis B	.20	.22	.13
Duration Target Spacing Analysis C	.24	.19	.25
Aircraft Distance Flown Under Control	-.04	-.05	.00
Fuel Consumption Under Control	-.03	-.05	.02
Arrival Altitudes Attained Completed Flights	-.03	.16	-.06
Time in Boundary	.08	-.01	-.03

TABLE 5. OCCUPANCY FACTOR

Factor Score Coefficients

	Day 1	Day 2	Day 3
Path Changes	.01	-.15	.00
Handoff Accept Delay Time	-.18	.20	.07
Number of Ground to Air Contacts	-.03	-.04	-.02
Duration of Ground to Air Contacts	-.05	.03	-.00
Total Delays	.00	-.06	.11
Total Delay Time	.06	-.06	-.13
Time Under Control	.28	.39	.32
Target Spacing Analysis A	.03	-.05	.01
Target Spacing Analysis B	.04	-.03	.02
Target Spacing Analysis C	-.03	-.04	-.06
Duration Target Spacing Analysis A	-.03	.08	-.00
Duration Target Spacing Analysis B	.03	.08	.10
Duration Target Spacing Analysis C	-.06	.00	-.04
Aircraft Distance Flown Under Control	.27	.03	.16
Fuel Consumption Under Control	.28	.25	.25
Arrival Altitudes Attained Completed Flights	-.04	-.28	-.02
Time in Boundary	.15	.46	.34

TABLE 6. COMMUNICATIONS FACTOR

Factor Score Coefficients

	Day 1	Day 2	Day 3
Path Changes	.34	.37	.36
Handoff Accept Delay Time	.06	-.04	.00
Number of Ground to Air Contacts	.38	.36	.37
Duration of Ground to Air Contacts	.39	.33	.38
Total Delays	-.02	.03	-.05
Total Delay Time	-.05	.00	.02
Time Under Control	-.01	-.04	-.01
Target Spacing Analysis A	-.07	-.02	.00
Target Spacing Analysis B	-.11	.01	-.06
Target Spacing Analysis C	-.00	.01	.02
Duration Target Spacing Analysis A	.09	-.03	-.02
Duration Target Spacing Analysis B	.01	.02	-.05
Duration Target Spacing Analysis C	.10	-.03	.06
Aircraft Distance Flown Under Control	-.05	-.04	-.02
Fuel Consumption Under Control	-.04	-.09	.02
Arrival Altitudes Attained Completed Flights	.04	.23	.03
Time in Boundary	.09	.00	-.01

TABLE 7. DELAY FACTOR

Factor Score Coefficients

	Day 1	Day 2	Day 3
Path Changes	-.06	-.00	.07
Handoff Accept Delay Time	.20	.06	.09
Number of Ground to Air Contacts	-.03	-.01	-.01
Duration of Ground to Air Contacts	-.03	.01	-.09
Total Delays	.43	.43	.53
Total Delay Time	.45	.43	.55
Time Under Control	.03	-.07	-.04
Target Spacing Analysis A	.04	-.07	-.06
Target Spacing Analysis B	.16	-.15	-.10
Target Spacing Analysis C	.00	.10	-.01
Duration Target Spacing Analysis A	-.13	.07	.10
Duration Target Spacing Analysis B	-.08	-.10	-.05
Duration Target Spacing Analysis C	-.11	.25	-.00
Aircraft Distance Flown Under Control	.08	.02	.05
Fuel Consumption Under Control	-.01	.04	.03
Arrival Altitudes Attained Completed Flights	.05	-.05	.01
Time in Boundary	-.12	-.07	-.13

TABLE 8. FACTOR SCORE CROSS VALIDATION CORRELATION (SHEET 1 of 2)

Conflict Factor

Factor Scores Computed Using Standard Scores of Day One Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.9337	.9802
Day 2 Coefficients		1.0000	.9259
Day 3 Coefficients			1.0000

Factor Scores Computed Using Standard Scores of Day Two Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.9377	.9752
Day 2 Coefficients		1.0000	.9293
Day 3 Coefficients			1.0000

Factor Scores Computed Using Standard Scores of Day Three Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.9272	.9769
Day 2 Coefficients		1.0000	.9261
Day 3 Coefficients			1.0000

Throughput Factor

Factor Scores Computed Using Standard Scores of Day One Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.8146	.9438
Day 2 Coefficients		1.0000	.9141
Day 3 Coefficients			1.0000

Factor Scores Computed Using Standard Scores of Day Two Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.6924	.8797
Day 2 Coefficients		1.0000	.8991
Day 3 Coefficients			1.0000

Factor Scores Computed Using Standard Scores of Day Three Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.7401	.8954
Day 2 Coefficients		1.0000	.9184
Day 3 Coefficients			1.0000

TABLE 8. FACTOR SCORE CROSS VALIDATION CORRELATION (SHEET 2 of 2)

Communications Factor

Factor Scores Computed Using Standard Scores of Day One Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.9282	.9789
Day 2 Coefficients		1.0000	.9425
Day 3 Coefficients			1.0000

Factor Scores Computed Using Standard Scores of Day Two Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.9429	.9855
Day 2 Coefficients		1.0000	.9535
Day 3 Coefficients			1.0000

Factor Scores Computed Using Standard Scores of Day Three Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.9316	.9802
Day 2 Coefficients		1.0000	.9559
Day 3 Coefficients			1.0000

Delay Factor

Factor Scores Computed Using Standard Scores of Day One Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.8462	.9404
Day 2 Coefficients		1.0000	.9411
Day 3 Coefficients			1.0000

Factor Scores Computed Using Standard Scores of Day Two Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.8650	.9440
Day 2 Coefficients		1.0000	.9610
Day 3 Coefficients			1.0000

Factor Scores Computed Using Standard Scores of Day Three Data

	Day 1 Coefficients	Day 2 Coefficients	Day 3 Coefficients
Day 1 Coefficients	1.0000	.8096	.9329
Day 2 Coefficients		1.0000	.9251
Day 3 Coefficients			1.0000

TABLE 9. FACTOR SCORE COEFFICIENTS FOR FULL FACTORS

Factor Analysis of SEM II Data

		Conflict	Occupancy	Communication	Delay
DA01	Path Changes	-.02	.00	.36	.00
DA09	Hand-off Accept Delay Time	.00	.07	.00	.09
DA10	Number Ground-to-Air Contacts	.01	-.03	.37	-.01
DA11	Duration Ground-to-Air Contacts	.01	.00	.38	-.03
DA12	Total Delays	-.06	.00	-.02	.43
DA13	Total Delay Time	-.04	-.06	.00	.45
DB02	Time Under Control	.01	.32	-.01	-.04
DB04	TSA-4 (Number of 4 Mile Conflicts)	.19	.03	-.02	-.06
DB05	TSA-5 (Number of 5 Mile Conflicts)	.18	.02	-.06	-.10
DB06	TSA-3 (Number of 3 Mile Conflicts)	.19	-.04	.01	.00
DB07	Duration TSA-4 (Duration of 4 Mile Conflicts)	.20	-.03	-.02	.07
DB08	Duration TSA-5 (Duration of 5 Mile Conflicts)	.20	.08	.01	-.08
DB09	Duration TSA-3 (Duration of 3 Mile Conflicts)	.24	-.04	.06	.00
DB10	Aircraft Distance Flown Under Control	.00	.16	-.04	.05
DB11	Fuel Consumption Under Control	-.03	.25	-.04	.03
DB13	Arrival Altitude Attained Completed Flights	-.03	-.02	.04	.01
DB15	Time in Boundary	-.01	.34	.00	-.12

TABLE 10. FACTOR SCORE COEFFICIENTS FOR VERY SMOOTH FACTORS

Factor Analysis of SEM II Data

		Conflict	Occupancy	Communication	Delay
DA01	Path Changes	*	*	.36	*
DA09	Hand-off Accept Delay Time	*	*	*	.09
DA10	Number Ground-to-Air Contacts	*	*	.37	*
DA11	Duration Ground-to-Air Contacts	*	*	.38	*
DA12	Total Delays	*	*	*	.43
DA13	Total Delay Time	*	*	*	.45
DB02	Time Under Control	*	.32	*	*
DB04	TSA-4 (Number of 4 Mile Conflicts)	.19	*	*	*
DB05	TSA-5 (Number of 5 Mile Conflicts)	.18	*	*	*
DB06	TSA-3 (Number of 3 Mile Conflicts)	.19	*	*	*
DB07	Duration TSA-4 (Duration of 4 Mile Conflicts)	.20	*	*	*
DB08	Duration TSA-5 (Duration of 5 Mile Conflicts)	.20	*	*	*
DB09	Duration TSA-3 (Duration of 3 Mile Conflicts)	.24	*	*	*
DB10	Aircraft Distance Flown Under Control	*	.16	*	*
DB11	Fuel Consumption Under Control	*	.25	*	*
DB13	Arrival Altitude Attained Completed Flights	*	*	*	*
DB15	Time in Boundary	*	.34	*	*

(* = .00)

TABLE 11. FACTOR SCORE COEFFICIENTS FOR VERY SMOOTH FACTORS

Factor Analysis of SEM II Data

		Conflict	Occupancy	Communication	Delay
DA01	Path Changes	*	*	.37	*
DA09	Hand-off Accept Delay Time	*	*	*	*
DA10	Number Ground-to-Air Contacts	*	*	.37	*
DA11	Duration Ground-to-Air Contacts	*	*	.37	*
DA12	Total Delays	*	*	*	.44
DA13	Total Delay Time	*	*	*	.44
DB02	Time Under Control	*	.26	*	*
DB04	TSA-4 (Number of 4 Mile Conflicts)	.20	*	*	*
DB05	TSA-5 (Number of 5 Mile Conflicts)	.20	*	*	*
DB06	TSA-3 (Number of 3 Mile Conflicts)	.20	*	*	*
DB07	Duration TSA-4 (Duration of 4 Mile Conflicts)	.20	*	*	*
DB08	Duration TSA-5 (Duration of 5 Mile Conflicts)	.20	*	*	*
DB09	Duration TSA-3 (Duration of 3 Mile Conflicts)	.20	*	*	*
DB10	Aircraft Distance Flown Under Control	*	.26	*	*
DB11	Fuel Consumption Under Control	*	.26	*	*
DB13	Arrival Altitude Attained Completed Flights	*	*	*	*
DB15	Time in Boundary	*	.26	*	*

(* = .00)

TABLE 12. SORTED ROTATED FACTOR LOADING, SEM I, SECTOR 14 DENSITY 1

	FACTOR	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
DB02	15	.988	.000	.000	.000
DB10	23	.975	.000	.000	.000
DB11	24	.971	.000	.000	.000
DB15	28	.604	.564	.000	.000
DA13	13	.000	.800	.000	.000
DB05	18	.000	.757	.000	.000
DB04	17	.000	.695	.330	.000
DB08	21	.378	.693	.320	.000
DA12	12	-.251	.679	.268	.000
DB06	19	.000	.000	.953	.000
DB09	22	.000	.000	.937	.000
DB07	20	.000	.296	.824	.000
DA11	11	.000	.000	.000	.815
DA10	10	.000	-.262	.000	.802
DA01	1	.000	.000	.000	.795
DA09	9	.251	.000	.000	.698
DB13	26	.000	.000	.259	.337
	VP	3.596	3.283	2.888	2.702

The above factor loading matrix has been rearranged so that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than 0.500 appear first, loadings less than 0.250 have been replaced by zero. For explanation of the abbreviations used in the first column, see appendix B.

TABLE 13. SORTED ROTATED FACTOR LOADING, SEM I, SECTOR 14 DENSITY 2

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
DB06	19	.934	.000	.000	.000
DB04	17	.929	.000	.000	.000
DB07	20	.922	.000	.000	.000
DB08	21	.914	.318	.000	.000
DB09	22	.911	.000	.000	.000
DB05	18	.850	.000	.000	.000
DA13	13	.822	.000	.255	-.300
DB02	15	.358	.909	.000	.000
DB10	23	.347	.895	.000	.000
DB11	24	.372	.886	.000	.000
DA12	12	.000	-.589	.000	-.258
DA10	10	.000	.000	.917	.000
DA01	1	.000	.000	.751	.000
DA11	11	.000	.254	.747	.000
DB13	26	.000	.000	.000	.866
DB15	28	.387	.000	.000	.820
DA09	9	.310	.000	.441	-.351
	VP	6.357	3.108	2.336	1.865

The above factor loading matrix has been rearranged so that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than 0.500 appear first, loadings less than 0.250 have been replaced by zero. For explanation of the abbreviations used in the first column, see appendix B.

TABLE 14. SORTED ROTATED FACTOR LOADING, SEM I, SECTOR 14 DENSITY 3

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
DB04	17	.936	.000	.000	.000
DB06	19	.922	.000	.000	.000
DB05	18	.906	.000	.000	.000
DB07	20	.887	.347	.000	.000
DB08	21	.873	.341	.000	.000
DB09	22	.801	.407	.000	.000
DB02	15	.356	.883	.000	.000
DB10	23	.365	.876	.000	.000
DB11	24	.362	.859	-.289	.000
DA09	9	.000	.527	.356	.264
DA12	12	.000	-.291	.878	.000
DA13	13	.000	.000	.861	.000
DB15	28	.421	.000	-.707	.338
DA10	10	.000	.000	.000	.853
DB13	26	.268	.000	.000	.677
DA01	1	.000	.000	.000	.660
DA11	11	.000	.356	.490	.615
	VP	5.425	3.278	2.735	2.238

The above factor loading matrix has been rearranged to that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than 0.500 appear first, loadings less than 0.250 have been replaced by zero. For explanation of the abbreviations used in the first column, see appendix B.

TABLE 15. SORTED ROTATED FACTOR LOADING, SEM I, SECTOR 16 DENSITY 1

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
DB04	17	.916	.000	.000	.000
DB08	21	.876	.000	.000	.000
DB06	19	.858	.000	.000	.000
DB07	20	.764	.295	-.432	.000
DB05	18	.713	-.387	.256	.000
DB10	23	.000	.967	.000	.000
DB02	15	.000	.952	.000	.000
DB11	24	.000	.948	.000	.000
DA10	10	.000	.000	.757	.000
DA01	1	.000	.436	.753	.000
DB09	22	.528	.306	-.663	.000
DA11	11	.000	.000	.627	.330
DB13	26	.000	.000	.000	.939
DA13	13	.000	.000	.000	-.836
DB15	28	.000	.508	.000	.640
DA09	9	-.391	.000	.000	.350
DA12	12	.000	.000	-.298	.000
	VP	3.934	3.665	2.481	2.277

The above factor loading matrix has been rearranged to that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than 0.500 appear first, loadings less than 0.250 have been replaced by zero. For explanation of the abbreviations used in the first column, see appendix B.

TABLE 16. SORTED ROTATED FACTOR LOADING, SEM I, SECTOR 14 DENSITY 2

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
DB11	24	.961	.000	.000	.000
DB02	15	.949	.000	.000	.000
DB10	23	.938	.000	.000	.000
DA12	12	-.692	.000	.000	.000
DA01	1	.542	.000	.486	.000
DB06	19	.000	.874	.000	.000
DB07	20	.000	.839	.000	.000
DB04	17	.000	.817	.000	.000
DB09	22	.000	.776	.000	.000
DA10	10	.000	.000	.852	.000
DA11	11	.000	.000	.836	.000
DA09	9	.000	.000	.806	.000
DB05	14	.000	.505	-.537	.000
DB13	26	.000	.000	.266	.897
DA13	13	.000	.000	-.264	-.839
DB15	28	.000	.000	.000	.741
DB08	21	.284	.416	-.409	-.374
	VP	3.686	3.305	3.031	2.432

The above factor loading matrix has been rearranged to that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than 0.500 appear first, loadings less than 0.250 have been replaced by zero. For explanation of the abbreviations used in the first column, see appendix B.

TABLE 17. SORTED ROTATED FACTOR LOADING, SEM I, SECTOR 16 DENSITY 3

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
DB07	20	.931	.000	.000	.000
DB09	22	.926	.000	.000	.000
DB05	18	.911	.000	.000	.000
DB04	17	.909	.000	.000	.000
DB06	19	.894	.000	.000	.000
DB08	21	.803	.340	-.258	.000
DB11	24	.000	.951	.000	.000
DB10	23	.000	.947	.000	.000
DB02	15	.000	.928	.000	.000
DA12	12	.000	-.623	.317	.000
DA09	9	.000	.000	.787	.000
DA11	11	-.364	.000	.718	.000
DA10	10	-.380	.000	.617	.000
DA01	1	.000	.386	.507	.000
DA13	13	.000	.000	.000	-.869
DB13	26	.000	.000	.000	.850
DB15	28	-.255	.446	.000	.536
	VP	5.361	3.718	2.051	1.984

The above factor loading matrix has been rearranged to that the columns appear in decreasing order of variance explained by factors. The rows have been rearranged so that for each successive factor, loadings greater than 0.500 appear first, loadings less than 0.250 have been replaced by zero. For explanation of the abbreviations used in the first column, see appendix B.

APPENDIX E

COMPUTATIONS OF RUN SCORES BASED ON THE INDEX OF ORDERLINESS

THE INDEX OF ORDERLINESS

A. George Halverson derived the index of orderliness as a measure of the risk of collision of an air traffic control situation. References 10, 11, 12, and 13 of the main body of this report contain the technical background. Halverson's original work, particularly as described in an unpublished technical note of August 1971, "Index of Orderliness: Proposed Measure of ATC System Performance", contains many alternative formulations. Some of these allow for accelerated motion, turns, etc. In many cases the index values are not constrained to lie between zero and one. Some of these indices are inversely proportional to the miss circle or miss volume with or without a time-dependent exponential damping term. Halverson discussed several means of obtaining an overall rating, including frequency analyses and use of autocorrelation functions.

In the Air Traffic Control Simulation Facility (ATCSF) the instantaneous index of orderliness for two targets has been implemented in the form:

(1) $ADD = z_n \cdot r_n \cdot e^{-tm}$

where: t_m is the time to minimum horizontal separation, in minutes
r_n is the normalized horizontal separation at minimum horizontal separation (CPA)

and: z_n is the normalized vertical separation at CPA.

This version of the index of orderliness is essentially a measure of the risk (probability) of a confliction occurring if no control action is taken and all targets continue on straight, unaccelerated flight paths. The index is roughly proportional to the ratio of a) the volume of a cylinder with height equal to the altitude separation at CPA (ZRMIN) and radius of the horizontal miss distance (RMIN) to b) the volume of a cylinder of height equal to the critical altitude separation, ZRCR, and radius of the critical horizontal separation, RCR. The negative exponential term discounts potential conflictions in terms of their distance in time.

In the ATCSF the value of the risk index, ADD, is calculated every simulation time step (normally every second) for all active targets, pairwise. The calculations performed during data reduction and analysis (DR&A) in the ATCSF are as follows:

Consider two targets (1) and (2), with coordinates (x(1),y(1),z(1)) and (x(2), y(2),z(2)). Define their respective velocity components as (XDOT(1),YDOT(1), ZRATE(1)) and (XDOT(2),YDOT(2),ZRATE(2)).

Then:
Separation between the two targets is -
in X coordinate, $XR=X(1)-X(2)$

in Y coordinate, $YR=Y(1)-Y(2)$

in Z coordinate, $ZR=Z(1)-Z(2)$

and the square of the horizontal separation, $RSQ = XR^2+YR^2$

Relative velocity components-
in X, $XRDOT = XDOT(1)-XDOT(2)$

in Y, $YRDOT = YDOT(1)-YDOT(2)$

in Z, $ZRATER = ZRATE(1)-ZRATE(2)$

Horizontal speed, $SPEED = (XRDOT^2+YRDOT^2)^{\frac{1}{2}}$

Relative distance to CPA, $PATHL = XR \cdot XRDOT+YR \cdot YRDOT$

Horizontal separation at closest point of approach (CPA) -
$RMIN = |YRDOT \cdot XR - XRDOT \cdot YR| /SPEED$

Time to CPA -
$TRMIN = -PATHL/(SPEED^2)$

Vertical separation at CPA -
$ZRMIN = |ZR+ZRATE \cdot TRMIN|$

For SEM the critical horizontal separation, RCR, was set at 10.0 nmi, and the critical vertical separation, ZRCR, was set at 1,000 feet.

Equation (1) becomes:

$$(2) \quad ADD = e^{-(TRMIN/60)} \cdot \frac{(ZRCR-ZRMIN)}{ZRCR} \cdot \frac{(RCR^2-RMIN^2)}{RCR^2}$$

where ADD is the instantaneous index for two targets.

The instantaneous (or every second) risk index, ADD, was subjected to a set of constraints. ADD was set to zero if-

a) ADD calculated is less than 0.01

b) Range is not closing, i.e.: $RSQ_{t=i} \geq RSQ_{t=i-1}$

c) The minimum range at CPA, RMIN, is greater than RCR

d) The minimum altitude separation at CPA, ZRMIN, is greater than ZRCR

e) Time to CPA,TRMIN, is greater than 420 seconds (7 minutes)

f) Either target is a departure flying below 1,000 feet.

g) Targets are locked onto parallel ILS courses

h) Either target has landed or is inactive (during the one-minute interval)

The risk measure for a pair of targets for a minute is taken as the maximum value of ADD for that pair for that minute. The risk for a controller (subject) for a minute is the risk of at least one confliction occurring during that minute. This is equal to 1.0 less the risk of no conflictions, which is the product over the pairs of 1.0 minus the risk of confliction.

$$(3)\quad I00=1.-(1.-ADD_1)\cdot(1.-ADD_2)\cdot(1.-ADD_3)\cdots(1.-ADD_n)=1.-\prod_{i=i}^{i=n}(1.-ADD_i)$$

A single value is needed to express the index of orderliness for a run. Three different cumulation methods were evaluated for obtaining a measure compatable with the SEM measure set and the SEM experimental conditions. These were the arithmetic mean,ORD1, the variance, ORD2, and the cumulative probability function,ORD3, of the index.

For a run of n minutes duration, the minute-by-minute values of the index,I00, (equation 3) are cumulated by:

$$(4)\quad ORD1=\frac{\sum_{i=1}^{i=n} I00_i}{n}=\overline{I00}$$

$$(5)\quad ORD2=\frac{\sum_{i=1}^{i=n}(I00_i-\overline{I00})^2}{(n-1)}$$

$$(6)\quad ORD3=1.0-\prod_{i=1}^{i=n}(1.-I00_i)$$

Note that ORD3 will be identically 1.0 if at any instant during the simulation the risk of confliction is 1.0. In addition, the maximum value of ORD3 would be 1.0, no matter what else occurred in the balance of the run.

APPENDIX F

LIST OF TERMINAL AREA SYSTEM EFFECTIVENESS MEASURES

Given below is a list of proposed measures for SEM experiments in the terminal environment. The major feature of these measures is their division into groups as follows.

Group A - System measures (Delays, Throughput, Communications)

Group B - System measures (Conflicts)

Group C - Radar Advisory Aircraft

Group D - IFR Aircraft

Group E - VFR Aircraft

All data measures will be calculated for the controller team as well as the North and South controllers individually.

Group A - (System Elements)

1. Number of Aircraft Handled - The number of aircraft entering the boundary of the sector, defined as being within the sectors vertical and horizontal limits (10,000 feet by 38 nautical miles from the radar center.

2. Number of Completed Flights - Flights entering the boundary and reaching ultimate points; arrivals - the middle marker; departures - the system boundary (horiz. or vert.) at or above a specified altitude, over or within 5 miles of a specified fix. A fix passage plus or minus 5 miles will be sensed even though passage may be well above the sector horizontal boundary.

Altitudes -	IFR Types 1-8	> 3000 ft.
	IFR Types 9-12	> 6000 ft.
	VFR All	> 2500 ft.

3. <u>Aircraft Time Under Control</u> - The amount of time aircraft are within the boundary, summed over all aircraft

4. <u>Number of Start Delays to Aircraft</u> - The number of instances that an aircraft was scheduled to enter the problem while a STOP message was in effect.

5. <u>Turn and Hold Delays</u> - The number of occasions aircraft within the boundary are put into a hold or a turn lasting more than 70 seconds.

6. <u>Total Delays</u> - Turn and Hold Delays plus Start Delays.

7. <u>Start Delay Duration</u> - The cumulative duration of Start Delays. For each affected aircraft, the start delay equals the difference between its scheduled start time and the time a start message is entered.

8. <u>Turn and Hold Duration</u> - The cumulative duration of Turn and Hold Delays within the boundary.

9. <u>Total Delay Duration</u> - The cumulative duration of Start Delays as well as Hold and Turn Delays within the boundary.

10. <u>Number of Path Changes</u> - The number of altitude, heading, and speed changes issued to aircraft within the boundary.

11. <u>Number of Path Changes Outside Boundary</u> - The number of altitude, heading, and speed changes issued to aircraft outside the boundary.

12. <u>Number of Handoffs Accepted</u> - The total number of aircraft handed off and accepted by the subject controller (inside the boundary, outside the boundary, and north to south within the boundary).

13. <u>Hand-off Accept Delay Time</u> - The cumulative time between a handoff and the acceptance of that aircraft by the subject controller.

14. <u>Number of Handoffs Outside the Boundary</u> - The total number of aircraft handed off and accepted by the subject controller outside the boundary.

15. North-South Hand-offs Accepted - The total number of aircraft handed off between the two members of the controller team.

16. North-South Hand-off Delay Time - The cumulative duration of North-South Hand-offs Accepted.

17. Aircraft Distance Flown - The distance flown by aircraft within the boundary summed over all aircraft.

18. Aircraft Fuel Consumption - The cumulative fuel in pounds consumed by aircraft within the boundary computed using the ATCSF fuel consumption model.

19. Number of Arrivals - The number of completed arrivals for both IFR and VFR aircraft.

20. Number of Departures - The number of departures for both IFR and VFR aircraft.

21. Departure Altitude Not Attained - The number of departing aircraft which do not climb above:

    IFR (Category 1-8) - 3000 feet

    IFR (Category 9-12) - 6000 feet

    VFR - 2500 feet

22. Missed Approaches - The number of system generated missed approaches. Aircraft misaligned with the ILS are spontaneously sent into missed approach status.

23. Ground-to-Air Contacts - The number of times microphone transmission is made by the subject or team.

24. Ground-to-Air Contacts Duration - The cumulative time of ground-to-air contacts.

25. Arrival Interval (Seconds) - The average number of seconds between completed arrivals.

26. Arrival Interval Variance (Seconds) - The variance in the distribution of arrival intervals.

27. Arrival Interval (Miles x 100) - The average number of miles between an arrival and the next arrival for all arrivals in the 60 minute test period times 100.

28. Arrival Interval Variance (Miles x 100) - The variance in the distribution of Arrival Intervals for miles x 100.

29. ILS Clearances - The number of aircraft cleared to the Instrument Landing System (ILS).

30. Control Actions After ILS Approach Clearance - Aircraft cleared for ILS approach will complete that approach unless another clearance, other than a speed control is given. These actions, after the approach clearance, are counted and shown under this heading.

    Missed approaches: The ATCSF already provides an automatic missed approach if an aircraft which has been cleared for an ILS approach is physically positioned such that it is impossible to perform the approach. The controller has the option of requiring vectors for spacing after an approach clearance.

31. Number of Barrier Delays - The number of instances a subject asks that all entering traffic be halted.

32. Barrier Delay Duration - The cumulative time that barrier delays remain in effect. The beginning of a barrier delay is referred to as a STOP message and its termination as a START message.

33. Aircraft Displayed - The total number of aircraft displayed on the CRT.

34. Aircraft Time Displayed - The cumulative duration of time in which active aircraft are displayed regardless of their position or classification.

35. Total Fuel Consumption - The cumulative fuel consumption of all active aircraft in the problem regardless of their position or classification.

36. Total Distance Flown - The cumulative distance flown by all active aircraft in the problem regardless of their position or classification.

37. Uncontrolled Aircraft Displayed - The number of uncontrolled aircraft displayed.

38. Uncontrolled Aircraft Time Displayed - The cumulative duration in which uncontrolled aircraft are displayed.

39. Controller Keyboard Errors - Keyboard errors which are detectable as such through the baseline ATCSF software.

40. Pilot Keyboard Errors - Keyboard errors by simulator operators which are detectable as such through the baseline ATCSF software.

Group B - (System Elements)

41. Target Spacing Analysis 4.0 for IFR Aircraft (TSIFR 4.0-950 ft.) - The number of instances that IFR aircraft violate the separation standard of 4 miles horizontal spacing and 950 feet vertical spacing. Both aircraft involved must be under IFR control and within the boundary.

42. Target Spacing Analysis 3.0 for IFR Aircraft (TSIFR 3.0-950 ft.) - Same as TSIFR 4.0 except horizontal separation is 3 miles.

43. Target Spacing Analysis 2.5 for IFR Aircraft (TSIFR 2.5-950 ft.) - Same as TSIFR 3.0 except horizontal separation is 2.5 miles.

44. Target Spacing Analysis 2.0 for IFR Aircraft (TSIFR 2.0-950 ft.) - Same as TSIFR 2.5 except horizontal separation is 2.0 miles.

45. Target Spacing Analysis 1.0 for IFR Aircraft (TSIFR 1.0-950 ft.) - Same as TSIFR 2.0 except horizontal separation is 1.0 mile

46. Duration TSIFR 4.0 - The cumulative duration of 4.0 mile conflicts for IFR aircraft.

47. Duration TSIFR 3.0 - The cumulative duration of 3.0 mile conflicts for IFR aircraft.

48. Duration TSIFR 2.5 - The cumulative duration of 2.5 mile conflicts for IFR aircraft.

49. Duration TSIFR 2.0 - The cumulative duration of 2.0 mile conflicts for IFR aircraft.

50. Duration TSIFR 1.0 - The cumulative duration of 1.0 mile conflicts for IFR aircraft.

51. Target Spacing Analysis 2.0 for VFR Aircraft (TSVFR 2.0-450 ft.) - The number of instances that VFR aircraft violate the separation standard of 2.0 miles horizontal spacing and 450 ft. vertical spacing below a height of 6,500 feet and within a radius of 10 miles of the radar center. At least one aircraft must be VFR.

52. Target Spacing Analysis 1.5 for VFR Aircraft (TSVFR 1.5-450 ft.) - Same as TSVFR 2.0, but with horizontal separation of 2.0 miles.

53. Target Spacing Analysis 1.0 for VFR Aircraft (TSVFR 1.0-450 ft.) - Same as TSVFR 1.5, but with horizontal separation of 1.0 mile.

54. Duration TSVFR 2.0 - The cumulative duration of 2.0 mile conflicts for VFR aircraft.

55. Duration TSVFR 1.5 - The cumulative duration of 1.5 mile conflicts for VFR aircraft.

56. Duration TSVFR 1.0 - The cumulative duration of 1.0 mile conflicts for VFR aircraft.

57. Target Spacing Analysis 6.0 for Aircraft on the ILS (TSILS 6.0) - The number of instances that appropriate categories of aircraft violate the 6.0 mile separation standard in the table below.

Conflict Separation Parameters

Index No.	Trailing A/C Size	Lead A/C Size	Horizontal Separation (Pillbox Radius)
1.	Small	Small	3 miles
2.	Small	Large	4 miles
3.	Small	Heavy	6 miles
4.	Large	Small	3 miles
5.	Large	Large	3 miles
6.	Large	Heavy	5 miles
7.	Heavy	Small	3 miles
8.	Heavy	Large	3 miles
9.	Heavy	Heavy	4 miles

58. Target Spacing Analysis 4.0 for Aircraft on the ILS (TSILS 4.0) - The same as TSILS 5.0, except separation is 4.0 miles.

59. Target Spacing Analysis 4.0 for Aircraft on the ILS (TSILS 4.0) - The same as TSILS 5.0, except horizontal separation is 4.0 miles.

60. Target Spacing Analysis 3.0 for Aircraft on the ILS (TSILS 3.0) - The same as TSILS 4.0, except horizontal separation is 3 0 miles.

61. Duration of TSILS 6.0 - The cumulative duration of 6.0 mile conflicts for aircraft on the ILS.

62. Duration of TSILS 5.0 - The cumulative duration of 5.0 mile conflicts for aircraft on the ILS.

63. Duration of TSILS 4.0 - Same as above, but for 4.0 mile conflicts.

64. Duration of TSILS 3.0 - Same as above, but for 3.0 mile conflicts.

65. ARTS Conflict Alert - The number of ARTS conflict alerts.

66. IFR (3 mile) Conflicts Outside Boundary - The number of three mile conflicts occurring outside the boundary for IFR aircraft.

Group C - (Radar Advisory Aircraft)

The list of measures below is defined here only for Radar Advisory Aircraft. The counts and durations of these measures are computed for Radar Advisory Aircraft only. In every other respect their definition is identical to the analogous system elements in Group A.

67. Number of Aircraft Handled (RA)

68. Aircraft Time Under Control (RA)

69. Number of Start Delays to Aircraft (RA)

70. Turn and Hold Delays (RA)

71. Total Delays (RA)

72. Start Delay Duration (RA)

73. Turn and Hold Duration (RA)

74. Total Delay Duration (RA)

75. Number of Path Changes (RA)

76. North-South Handoff Accepts (RA)

77. North-South Handoff Accept Delay Time (RA)

78. Aircraft Distance Flown (RA)

79. Aircraft Fuel Consumption (RA)

Group D - (IFR Aircraft)

The list of measures below is defined here only for IFR aircraft. The counts and durations of these measures are computed for IFR aircraft only. In every other respect their definition is identical to the analogous system elements in Group A.

80. Number of Aircraft Handled (IFR)

81. Number of Completed Flights (IFR)

82. Aircraft Time Under Control (IFR)

83. Number of Start Delays to Aircraft (IFR)

84. Turn and Hold Delays (IFR)

85. Total Delays (IFR)

86. Start Delay Duration (IFR)

87. Turn and Hold Duration (IFR)

88. Total Delay Duration (IFR)

89. Number of Path Changes (IFR)

90. Number of Handoffs Accepted (IFR)

91. Handoff Accept Delay Time (IFR)

92. North-South Handoff Accepts (IFR)

93. North-South Handoff Accept Delay Time (IFR)

94. Aircraft Distance Flown (IFR)

95. Aircraft Fuel Consumption (IFR)

96. Arrivals (IFR)

97. Departures (IFR)

98. Departure Altitude Not Attained (IFR)

99. Missed Approaches (IFR)

Group E - (VFR Aircraft)

The list of measures below is defined here only for VFR aircraft. The counts and durations of these measures are computed for VFR aircraft only. In every other respect their definition is identical to the analogous system elements in Group A.

100. Number of Aircraft Handled (VFR)

101. Number of Completed Flights (VFR)

102. Aircraft Time Under Control (VFR)

103. Number of Start Delays to Aircraft (VFR)

104. Turn and Hold Delays (VFR)

105. Total Delays (VFR)

106. Start Delays Duration (VFR)

107. Turn and Hold Duration (VFR)

108. Total Delay Duration (VFR)

109. Number of Path Changes (VFR)

110. North-South Handoff Accepts (VFR)

111. North-South Handoff Accept Delay Time (VFR)

112. Aircraft Distance Flown (VFR)

113. Aircraft Fuel Consumption (VFR)

114. Arrivals (VFR)

115. Departures (VFR)

116. Departure Altitude Not Attained (VFR)

☆U.S. GOVERNMENT PRINTING OFFICE: 1993 - 704-078/80059